UNIVERSIDAD CARLOS III DE MADRID working papers

UC3M Working Papers
Statistics and Econometrics
17-10
ISSN 2387-0303
Mayo 2017

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-48

Multivariate outlier detection based on a robust Mahalanobis distance with shrinkage estimators

Elisa Cabana[a], Henry Laniado[b], Rosa E. Lillo[c]

Abstract

A collection of methods for multivariate outlier detection based on a robust Mahalanobis distance is proposed. The procedure consists on different combinations of robust estimates for location and covariance matrix based on shrinkage. The performance of our proposal is illustrated, through the comparison to other techniques from the literature, in a simulation study. The resulting high correct classification rates and low false classification rates in the vast majority of cases, and also the good computational times shows the goodness of our proposal. The performance is also illustrated with a real dataset example and some conclusions are established.

Keywords: outlier detection, robust Mahalanobis distance, shrinkage estimator, high-dimension, robust estimation, robust location, robust covariance matrix.

[a] Department of Statistics, Universidad Carlos III de Madrid.
[b] Department of Mathematical Sciences, Universidad EAFIT, Medellín, Colombia.
[c] Department of Statistics and Institute of Financial Big Data, Universidad Carlos III de Madrid, Spain.

Acknowledgements:
This research was partially supported by Spanish Ministry grant ECO2015-66593-P.

1 Introduction

The detection of outliers in multivariate data is an important task in Statistics, since that kind of data can distort any statistical procedure. In data mining and machine learning contexts many standard techniques such as principal component analysis and linear discriminant analysis are inherently susceptible to atypical observations [Tarr et al., 2016]. The task of detecting multivariate outliers can be useful in various fields such as quality control, medicine, finance, image analysis, and chemistry (Vargas N [2003], Brettschneider et al. [2008], Hubert et al. [2008], Hubert and Debruyne [2010], Perrotta and Torti [2010] and Choi et al. [2016]). Outliers are generally defined as observations resulting from a secondary process, which differ from the background distribution. This kind of data do not need to be especially high or low in relation to all values of the variables in the data set, thus, this is the reason why the task of identifying multivariate outliers with the classical univariate methods commonly fail. In the multivariate case, there must be considered both the distance of an observation from the centroid of the data, and the shape of the data. The shape of multivariate data is characterized by the covariance matrix, and the Mahalanobis distance [Mahalanobis, 1936] is a well-known measure which takes it into account. For multivariate normally distributed data, the distribution of the squared Mahalanobis distance, MD^2, is known [Gnanadesikan and Kettenring, 1972] to be chi-squared with p (the dimension of the data) degrees of freedom, i.e. χ^2_p. Then, the adopted rule for identifying the outliers is selecting the threshold as the 0.975 quantile of the χ^2_p.

However, outliers need not necessarily have large MD values (Masking problem) and not all observations with large MD values are necessarily outliers (Swamping problem) [Hadi, 1992]. The problems of masking and swamping arise due to the influence of outliers on classical location and scatter estimates, i.e. the mean and the sample covariance matrix are not robust estimates, which implies that the estimated distance measures will not be reliable. The solution, in order to define a robust Mahalanobis distance measure RMD, is to consider robust estimates of centrality and covariance matrix. Rousseeuw [1985] proposed the Minimum Covariance Determinant (MCD) estimator based on the computation of the ellipsoid with the smallest volume or with the smallest covariance determinant that would encompass at least half of the data points. Consistency and asymptotic normality of the MCD estimator has been shown by Cator and Lopuhaä [2010] and Cator et al. [2012]. The procedure required naive subsampling for minimizing the objective function of the MCD, but an improvement much more effective, the Fast-MCD, was introduced by Rousseeuw and Driessen [1999] and a code is available in MATLAB (Verboven and Hubert [2005]). Unfortunately Fast-MCD still requires substantial running times for large p, because the number of candidate solutions grows exponentially with the dimension of the sample and, as a consequence, the procedure becomes computationally expensive for even moderately sized problems.

On the other hand, the squared RMD distributional fit usually breaks down, i.e. it does not necessary follow a chi-squared distribution when you apart from the Gaussian distribution. Thus, determining exact cutoff values for outlying distances continues to be a difficult problem and has found much attention because no universally applicable method has been proposed. Despite this fact, the $\chi^2_{p;0.975}$ quantile is often considered as threshold for recognizing outliers in the robust distance case, but this approach may have some drawbacks. Evidence of this behavior is now well documented even in moderately large samples, especially when the number of dimensions increases (Becker and Gather [1999], Hardin and Rocke [2005], Cerioli et al. [2009] and Cerioli et al. [2008]). It is crucial to determine the threshold for the distances in order to decide whether an observation is an outlier [Cerioli et al., 2009]. Related to the threshold selection problem, Filzmoser et al. [2005] proposed to use an adjusted quantile, instead of the classical choice of the $\chi^2_{p;0.975}$ quantile. The adjusted threshold is estimated adaptively from the data, but their proposal is defined for an specific robust Mahalanobis distance, the one based on the MCD estimators.

In Peña and Prieto [2001] and Peña and Prieto [2007] an algorithm is proposed, based on the analysis of the projections of the sample points onto a certain set of directions obtained by maximizing and minimizing the Kurtosis coefficient of the projections, and some random directions generated by a stratified sampling scheme. With the combination of random and specific directions the authors proposed a powerful procedure for robust estimation and outlier detection. The random directions could not completely cope with the deficiencies for concentrated contamination, but on the other hand, the specific directions obtained by the kurtosis coefficient seemed to be powerful for detecting concentrated contamination. However, this procedure has some drawbacks when the dimension of the sample space p grows.

Furthermore, there are several real scenarios where the number of variables is high in which outlier detection is very important. For example, medical imaging datasets often contain deviant observations due to acquisition or pre-processing artifacts or resulting from large intrinsic inter-subject variability. Specifically in Neuroimage, various kinds of acquisition artifacts may be present in fMRI (functional Magnetic Resonance Images) data, even small movements of the head may produce large artifacts in the signals, also heartbeat and breathing both induce pulsatile motion in the brain, which creates physiological noise artifacts directly in the data (Lazar [2008], Lindquist [2008], Monti [2011], Poline and Brett [2012]). High-dimensional data are increasingly encountered in other applications of statistics, e.g. in biological and financial studies (Chen et al. [2010] and Zeng et al. [2015]). Also, in geochemical data, because of their complex nature [Reimann and Filzmoser, 2000], regional geochemical datasets practically always contain outliers. In fact, finding data outliers that may be indicative of mineralization (in exploration geochemistry) or contamination (in environmental geochemistry) is one of the major aims of geochemical surveys [Templ et al., 2008].

In this article, a collection of methods based on a RMD are proposed. They are based on considering different combinations of robust estimates of location and covariance matrix. Two basic options are considered for the location parameter: a component-wise median and the L_1 multivariate median (Gower [1974], Brown [1983], Dodge [1987], Small [1990]). A notion called *Shrinkage estimator* (Ledoit and Wolf [2003a], Ledoit and Wolf [2003b], Ledoit and Wolf [2004], DeMiguel et al. [2013]) is considered, which is based on the fact that shrinking a sample estimator towards a target estimator would help to reduce the estimation error. The Shrinkage is applied to both of the previous mentioned location estimators, in some of the proposed combinations. As for the covariance matrix, the options basically consists on a Shrinkage estimator over special cases of *Comedian matrices* (Hall and Welsh [1985], Falk [1997]), as the sample estimator to base the shrinking on. In this paper, we analyzed the best option for shrinking the location and the scale parameters. By means of a simulations study, the satisfactory practical performance is shown, especially when the dimension of the problem grows. The computational cost is studied by both simulations and a real dataset example.

The paper is organized as follows. The description of the Robust Mahalanobis Distance derived from the mentioned procedures from the literature is presented in Section 2. Section 3 describes the Shrinkage estimators both for the location and the covariance matrix, and the proposed combinations of these estimator in order to define a RMD. Section 4 compares by simulations the proposed collection of RMD with the other approaches: the classical Mahalanobis distance, the RMD based on the MCD estimators with the classical quantile selected as threshold, the RMD based on the MCD estimators with the adjusted quantile selected as threshold, and the Kurtosis approach. It is shown that the proposed procedures have an advantageous behavior through all simulations specially when dimension increases. Section 5 shows the behavior of all methodologies with a real dataset example and it is shown that the proposed methods work well in practice and require reasonable computational times, even for large problems. Finally, Section 6 provides some conclusions.

2 The robust Mahalanobis distance

The classical Mahalanobis distance is defined for every $p-$dimensional observation $\mathbf{x_{i\cdot}}$ of the multivariate sample $\{\mathbf{x_{1\cdot}}, ..., \mathbf{x_{n\cdot}}\}$, as:

$$MD_i := ((\mathbf{x_{i\cdot}} - \hat{\boldsymbol{\mu}})\hat{\Sigma}^{-1}(\mathbf{x_{i\cdot}} - \hat{\boldsymbol{\mu}})^T)^{1/2}, \tag{1}$$

where $\hat{\boldsymbol{\mu}}$ is the estimated multivariate location (sample mean) and $\hat{\Sigma}$ is the estimated covariance matrix (sample covariance matrix).

The problem with this definition is that the classical estimates of location and covariance matrix used in the Mahalanobis distance equation (1) are often highly influenced by the presence of outliers [Rosseuw and Van Zomeren, 1990]. This fact means that the estimated distance measures will not be accurate. The solution is to consider robust estimates of centrality and covariance matrix, i.e. resistant against the influence of outlying observations, giving rise to a robust Mahalanobis distance, defined as:

$$RMD_i := ((\mathbf{x_{i\cdot}} - \hat{\boldsymbol{\mu}}_R)\hat{\Sigma}_R^{-1}(\mathbf{x_{i\cdot}} - \hat{\boldsymbol{\mu}}_R)^T)^{1/2}, \tag{2}$$

where $\hat{\boldsymbol{\mu}}_R$ and $\hat{\Sigma}_R$ are robust estimators of centrality and covariance matrix, respectively.

Many robust estimators for location and covariance have been introduced in the literature [Maronna and Yohai, 1976]. The most frequently used in practice, is the minimum covariance determinant (MCD) estimator [Rousseeuw, 1985]. This method consists on determining the subset J of observations of size h which minimizes the determinant of the sample covariance matrix, computed from only these h points. The choice of h determines the robustness of the estimator, in fact, it is a compromise between robustness and efficiency. Once this subset of size h is found, it is possible to estimate the centrality ($\hat{\boldsymbol{\mu}}_{MCD}$) and the covariance matrix ($\hat{\Sigma}_{MCD}$), based only upon that subset, and they will be robust estimates.

$$J = \left\{\text{set of } h \text{ points} : |\hat{\Sigma}_J| \leq |\hat{\Sigma}_K| \text{ for all subsets K s.t. } \#K = h\right\}$$

$$\hat{\boldsymbol{\mu}}_{MCD} = \frac{1}{h}\sum_{i\in J}\mathbf{x_{i\cdot}}$$

$$\hat{\Sigma}_{MCD} = \frac{1}{h}\sum_{i\in J}(\mathbf{x_{i\cdot}} - \hat{\boldsymbol{\mu}}_{MCD})(\mathbf{x_{i\cdot}} - \hat{\boldsymbol{\mu}}_{MCD})^T,$$

where $|A|$ denotes the determinant of the matrix A, and $\#K$ denotes the cardinality of the subset K.

Using the MCD robust estimators in the definition of the Mahalanobis distance, gives place to a robust measure:

$$RMD_{MCD}(\mathbf{x_{i\cdot}}) = ((\mathbf{x_{i\cdot}} - \hat{\boldsymbol{\mu}}_{MCD})\hat{\Sigma}_{MCD}^{-1}(\mathbf{x_{i\cdot}} - \hat{\boldsymbol{\mu}}_{MCD})^T)^{1/2} \tag{3}$$

The rule in this approach for detecting outliers is considering the classical threshold value as the $\chi^2_{p;0.975}$ quantile.

Fixing the threshold value ($\chi^2_{p;0.975}$) is rather subjective in the robust distance case, because:

1. There is no demonstration of the true distribution of the squared robust Mahalanobis distance.
2. There is no reason why this fixed threshold should be appropriate for every data set.
3. The threshold should be adjusted to the sample size [Reimann et al., 2005].
4. If the data come from a single multivariate normal distribution, there are no observations coming from a different distribution, only extremes. In this case, the threshold should be infinity.

Filzmoser et al. [2005] have stated that a better choice for selecting the threshold is to adjust it to the data set at hand. They proposed a method for estimating the threshold in the case of the robust Mahalanobis distance based on the MCD robust estimator, adaptively from the data. It basically consists on comparing the theoretical distribution function of χ^2_p and the empirical distribution function of the squared robust distance that has been chosen and finding the maximum positive difference between the two distributions, in the tail. This value serves as the alpha value for the threshold if the previous difference calculated does not exceed a critical value p_{crit}. The authors developed the critical value for the robust Mahalanobis Distance based on MCD estimators (3). In Section 4 it is shown that in most simulations scenarios the adjusted threshold improves the false classification rates, while maintaining the same correct classification rates, except in a few cases on which the correct classification rates will also be slightly declining.

Another approach is the one proposed in Peña and Prieto [2001] and Peña and Prieto [2007], which is based on the idea that very high or too low kurtosis coefficients suggest the presence of outliers. The authors take the projections of the sample points onto the set of directions obtained by maximizing and minimizing the Kurtosis coefficient and they consider also a set of random directions generated by a stratified sampling scheme. This is a powerful approach for robust estimation and outlier detection. However, when the dimension grows, the method worsens its performance and in presence of correlation between the variables the method looses power [Marcano and Fermín, 2013], being not very efficient computationally because of the optimization problem associated with the computation of the directions.

The authors proposed to project the "n" cloud of points in $\mathbb{R}^p$ over two new $p-$dimensional spaces: the first one obtained with the maximum kurtosis orthogonal direction, and the second one with the minimum kurtosis orthogonal direction, and also over a set of random directions. After obtaining the whole set of directions, the next step is to determine a "measure of outlyingness" for each observation (actually for their univariate projections $z_i^{(j)}$) as:

$$r_i = max_{1 \leq j \leq d} \frac{|z_i^{(j)} - median(\mathbf{z^{(j)}})|}{MAD(\mathbf{z^{(j)}})}, \tag{4}$$

where d is the total number of directions in which the data is projected, the univariate projections are $\mathbf{z^{(j)}} = (z_1^{(j)}, ..., z_n^{(j)})$, $median$ is the univariate median and MAD denotes the Median Absolute Deviation (Gauss [1816], Rousseeuw and Croux [1993], Leys et al. [2013]), which is a robust measure of the variability of a univariate sample and it is defined as the median of the absolute deviations from the data's median:

$$MAD(\mathbf{z^{(j)}}) = median \left(\left| z_i^{(j)} - median(\mathbf{z}^{(j)}) \right| \right)$$

With the above measure r_i they tested if a given observation is considered outlier if the condition r_i being greater than a certain cutoff value holds. If the condition holds for some i, a new sample composed of all observations whose r_i is less than the cutoff value is formed, and the procedure is applied again to the reduced sample. This is repeated until either no additional observations satisfy their r_i is greater than the cutoff value, or the number of remaining observations is less than $\lfloor (n + p + 1)/2 \rfloor$. Finally, a Mahalanobis distance is computed for all observations labeled as outliers in the preceding steps, using the mean and the covariance estimator based upon the remaining observations. Let U be the set of observations not labeled as outliers by the method, then the estimates of location $\hat{\boldsymbol{\mu}}_K$ and covariance matrix $\hat{\Sigma}_K$ (where the subscript K stands as a notation for "Kurtosis"), based upon this subset U defines a robust Mahalanobis distance as:

$$RMD_K(\mathbf{x_{i\cdot}}) = ((\mathbf{x_{i\cdot}} - \hat{\boldsymbol{\mu}}_K)^T \hat{\Sigma}_K^{-1} (\mathbf{x_{i\cdot}} - \hat{\boldsymbol{\mu}}_K))^{-1/2} \tag{5}$$

The final step is using this Mahalanobis distance to recover observations "mislabeled" as outliers, i.e. if the observation $i \notin U$ has $RMD_K(\mathbf{x_{i\cdot}}) < \chi^2_{p;0.99}$, then $\mathbf{x_{i\cdot}}$ is included in U. The process is repeated until no more such observations are found or U becomes the set of all observations.

3 A robust Mahalanobis distance based on Shrinkage estimator

There are several definitions for a robust Mahalanobis distance, depending on the robust estimators of centrality μ and covariance matrix Σ selected, as it is shown in the previous section. We propose to use a notion which is frequently used in Finance and Portfolio optimization, known as *Shrinkage*. Shrinkage estimators E_{Sh} consists on the fact that "shrinking" an estimator E of a parameter θ towards a *target estimator* T, would help to reduce the estimation error. Thus, the advantage is that although the shrinkage target is usually biased, it also contains less variance than the estimator E. Therefore, under general conditions, there exists a *shrinkage intensity* α, so the resulting shrinkage estimator would contain less estimation error than the estimator E [James and Stein, 1961]. In other words, a shrinkage estimator is a trade-off between a low bias estimator and a low variance estimator:

$$E_{Sh} = (1-\alpha)E + \alpha T \tag{6}$$

This approach can be applied both to the Location and Dispersion parameters needed for the robust Mahalanobis distance. In this paper, different combination of those parameters have been studied in detail.

3.1 Location parameter

For multivariate data $\mathbf{x} = \{\mathbf{x_{\cdot 1}}, ..., \mathbf{x_{\cdot p}}\}$, with each $\mathbf{x_{\cdot j}} \in \mathbb{R}^n$, for $j = 1, ..., p$, the classical choice for the Location and Dispersion parameters are the sample mean and the sample covariance matrix, respectively. But they are not robust to the presence of outliers as mentioned before. Based on the fact that the *median* is a better choice, in terms of robustness, we start by considering as a Location parameter the median that treats the columns of $\mathbf{x}$ as vectors and returns a row vector of median values which is called the *component-wise median:*

$$\hat{\boldsymbol{\mu}}_{R_1} = \hat{\boldsymbol{\mu}}_{CCM} = (median(\mathbf{x_{\cdot 1}}), ..., median(\mathbf{x_{\cdot p}})), \tag{7}$$

where *median* denotes the univariate median and $(\mathbf{x_{\cdot j}}) = (x_{1j}, ..., x_{nj})$ for all $j = 1, ..., p$.

Another option is to consider as the Location parameter a multivariate median $\hat{\boldsymbol{\mu}}_{MM}$ called $L_1-median$ (Gower [1974], Brown [1983], Dodge [1987], Small [1990]), which is a robust and highly efficient estimator of central tendency (Lopuhaa and Rousseeuw [1991], Vardi and Zhang [2000], Oja [2010]). It is defined as:

$$\hat{\boldsymbol{\mu}}_{R_2} = \hat{\boldsymbol{\mu}}_{MM} = argmin_{\mathbf{x_{m\cdot}}\ ,\ \mathbf{m} \in \{\mathbf{1},...,\mathbf{n}\}} \frac{1}{n} \sum_{i=1}^{n} |\mathbf{x_{m\cdot}} - \mathbf{x_{i\cdot}}| \tag{8}$$

In DeMiguel et al. [2013] a Shrinkage estimator over the sample mean is proposed, and in the same way we propose to study Shrinkage estimators for both of the above possibilities of Location, i.e. the component-wise median and the L_1-median.

Consider $\nu_{\boldsymbol{\mu}}\mathbf{e}$ as the target estimator T in (6), as in [DeMiguel et al., 2013], where $\mathbf{e}$ is the $p-$dimensional vector of ones, and consider $\hat{\boldsymbol{\mu}}_{CCM}$ as the sample estimator E. Then, the Shrinkage estimator over the component-wise median is:

$$\hat{\boldsymbol{\mu}}_{R_3} = \hat{\boldsymbol{\mu}}_{Sh(CCM)} = (1-\alpha)\hat{\boldsymbol{\mu}}_{CCM} + \alpha\nu_{\boldsymbol{\mu}}\mathbf{e} \tag{9}$$

The scaling factor $\nu_{\boldsymbol{\mu}}$ and the intensity α should minimize the bias of the shrinkage target, that is:

$$\begin{array}{cc} \min_{\nu_{\boldsymbol{\mu}},\alpha} & E\left[\left\|\hat{\boldsymbol{\mu}}_{Sh(CCM)} - \boldsymbol{\mu}\right\|_2^2\right] \\ \text{s.t.} & \hat{\boldsymbol{\mu}}_{Sh(CCM)} = (1-\alpha)\hat{\boldsymbol{\mu}}_{CCM} + \alpha\nu_{\boldsymbol{\mu}}\mathbf{e}, \end{array} \tag{10}$$

where $\|\mathbf{x}\|_2^2 = \sum_{j=1}^{p} x_j^2$.

The following result shows the choice of those parameters:

Proposition 1 *The scaling parameter $\nu_{\boldsymbol{\mu}}$ and the shrinkage intensity parameter α that minimizes the expected quadratic loss, given in (10), are:*

$$\nu_{\boldsymbol{\mu}} = \frac{\hat{\boldsymbol{\mu}}_{CCM}\mathbf{e}}{p} \tag{11}$$

$$\alpha = \frac{E\left[\|\hat{\boldsymbol{\mu}}_{CCM} - \boldsymbol{\mu}\|_2^2\right]}{E\left[\|\hat{\boldsymbol{\mu}}_{CCM} - \nu_{\boldsymbol{\mu}}\mathbf{e}\|_2^2\right]} \tag{12}$$

See the proof in Appendix A.

Note that the denominator for the α expression in the above expression (12) has no problem when estimating, but the numerator is not straightforward because $\boldsymbol{\mu}$ is unknown. Then it is necessary to note what follows:

Let $\mathbf{x} = \{\mathbf{x_{\cdot 1}}, ..., \mathbf{x_{\cdot j}}, ..., \mathbf{x_{\cdot p}}\}$. The component-wise median defined in (7) can also be described as:

$$\boldsymbol{\mu}_{CCM} = (\boldsymbol{\mu}_{CCM1}, ..., \boldsymbol{\mu}_{CCMj}, ..., \boldsymbol{\mu}_{CCMp}) \tag{13}$$

where $\boldsymbol{\mu}_{CCMj} = median(\mathbf{x_{\cdot j}})$.

In [Chu, 1955] the author investigated the distribution for the sample median estimator, and they obtained the following result about the variance in presence of normality. Fix j, for $j \in \{1, ..., p\}$:

$$\sigma^2_{\boldsymbol{\mu}_{CCMj}} = Var(\boldsymbol{\mu}_{CCMj}) = \frac{\pi}{2n}\sigma^2_{\mathbf{x_{\cdot j}}} \tag{14}$$

Therefore, the numerator in the expression (12) for determining the α in Proposition 1, is:

$$E\left[\|\hat{\boldsymbol{\mu}}_{CCM} - \boldsymbol{\mu}\|_2^2\right] = E\left[\sum_{j=1}^{p}(\hat{\boldsymbol{\mu}}_{CCMj} - \boldsymbol{\mu}_j)^2\right] = \sum_{j=1}^{p}\sigma^2_{\boldsymbol{\mu}_{CCMj}} = \frac{\pi}{2n}\sum_{j=1}^{p}\sigma^2_{\mathbf{x_{\cdot j}}} \tag{15}$$

Now, we propose to properly estimate each $\sigma^2_{\mathbf{x_{\cdot j}}}$ as explained in the next subsection (29).

On the other hand, consider $\nu_{\boldsymbol{\mu}}\mathbf{e}$ again as the target estimator T and consider $\hat{\boldsymbol{\mu}}_{MM}$ as the sample estimator E, in (6). Then, the Shrinkage estimator over the multivariate L_1-median is:

$$\hat{\boldsymbol{\mu}}_{R_4} = \hat{\boldsymbol{\mu}}_{Sh(MM)} = (1-\alpha)\hat{\boldsymbol{\mu}}_{MM} + \alpha\nu_{\boldsymbol{\mu}}\mathbf{e} \tag{16}$$

The scaling factor $\nu_{\boldsymbol{\mu}}$ and the intensity α should minimize the bias of the shrinkage target, as before:

$$\begin{array}{cc} \min_{\nu_{\boldsymbol{\mu}},\alpha} & E\left[\|\hat{\boldsymbol{\mu}}_{Sh(MM)} - \boldsymbol{\mu}\|_2^2\right] \\ \text{s.t.} & \hat{\boldsymbol{\mu}}_{Sh(MM)} = (1-\alpha)\hat{\boldsymbol{\mu}}_{MM} + \alpha\nu_{\boldsymbol{\mu}}\mathbf{e}, \end{array} \tag{17}$$

where $\|\mathbf{x}\|_2^2 = \sum_{j=1}^{p} x_j^2$.

Next proposition shows the optimal expression for the parameters $\nu_{\boldsymbol{\mu}}$ and α.

Proposition 2 *The scaling parameter $\nu_{\boldsymbol{\mu}}$ and the shrinkage intensity parameter α that minimizes the expected quadratic loss, given in (17), are:*

$$\nu_{\boldsymbol{\mu}} = \frac{\hat{\boldsymbol{\mu}}_{MM}\mathbf{e}}{p} \tag{18}$$

$$\alpha = \frac{E\left[\|\hat{\boldsymbol{\mu}}_{MM} - \boldsymbol{\mu}\|^2\right]}{E\left[\|\hat{\boldsymbol{\mu}}_{MM} - \nu_{\boldsymbol{\mu}}\mathbf{e}\|^2\right]} \tag{19}$$

See the proof in Appendix A.

As in the previous case, the denominator in the α expression (19) can be described as:

$$E\left[\|\hat{\boldsymbol{\mu}}_{MM} - \boldsymbol{\mu}\|_2^2\right] = E\left[\sum_{j=1}^{p}(\hat{\boldsymbol{\mu}}_{MMj} - \boldsymbol{\mu}_j)^2\right] = \sum_{j=1}^{p}\sigma^2_{\boldsymbol{\mu}_{MMj}}$$

In [Bose and Chaudhuri, 1993], [Bose, 1995] and [Möttönen et al., 2010], the authors investigated the asymptotic distribution for the L_1-median, and they obtained the following result about the covariance matrix in presence of normality:

$$\boldsymbol{\mu}_{MM} \sim N_p\left(\boldsymbol{\mu}, \frac{1}{n}\hat{A}^{-1}\hat{B}\hat{A}^{-1}\right), \tag{20}$$

where $A(\mathbf{x_{i\cdot}}) = \frac{1}{||\mathbf{x_{i\cdot}}||_2}\left(I_p - \frac{\mathbf{x_{i\cdot}}\mathbf{x_{i\cdot}}'}{||\mathbf{x_{i\cdot}}||_2^2}\right)$ and $B(\mathbf{x_{i\cdot}}) = \frac{\mathbf{x_{i\cdot}}\mathbf{x_{i\cdot}}'}{||\mathbf{x_{i\cdot}}||_2^2}$, with $\mathbf{x_{i\cdot}} \in \mathbb{R}^p$, for each $i = 1, ..., n$.

Then, the numerator in the expression (19) for determining the α in Proposition 2, is:

$$E\left[\|\hat{\boldsymbol{\mu}}_{MM} - \boldsymbol{\mu}\|_2^2\right] = trace\left(\frac{1}{n}\hat{A}^{-1}\hat{B}\hat{A}^{-1}\right) \tag{21}$$

3.2 Dispersion parameter

Based on the median concept, which is a robust measure of Location, one can define a robust measure of Dispersion which is the *Median Absolute Deviation (MAD)* from the data's median. Consider a random variable X:

$$MAD(X) = med(|X - med(X)|), \tag{22}$$

where med denotes median.

In [Falk, 1997], the following relation, assuming normality, between MAD and the standard deviation σ for a random variable X is given:

$$MAD(X) = \sigma_X \Phi^{-1}(3/4), \tag{23}$$

where Φ denotes the standard normal cdf.

Taking the square in (23) we obtain a relation, for a random variable X, between the variance σ_X^2 and $MAD^2(X)$:

$$\sigma_X^2 = 2.198 * MAD^2(X) \tag{24}$$

A robust measure of dependence between two variables arises in terms of the median. It is known as the *comedian* and it was introduced by [Falk, 1997]. Let X and Y be two random variables, then the *comedian* of X and Y is defined as:

$$COM(X, Y) = med((X - med(X))(Y - med(Y))) \tag{25}$$

The comedian generalizes the MAD, because $COM(X, X) = MAD^2(X)$ and also has the highest possible breakdown point [Falk, 1997].

An important fact is that $COM(X, Y)$ parallels $COV(X, Y)$, but the latter requires the existence of the first two moments of X and Y, whereas the comedian always exists. Other known properties of the comedian are that it is symmetric, location invariant and scale equivariant. Furthermore, Hall and Welsh [1985] discussed about the strong consistency and asymptotic normality of the MAD, and Falk [1997] established similar results for the comedian.

Finally, a comedian matrix can be defined based on a multivariate version of the comedian (25). Let $\mathbf{x} = \{\mathbf{x_{.1}}, ..., \mathbf{x_{.p}}\}$ be the $n \times p$ data matrix with n being the sample size and p the number of variables. Then the comedian matrix is defined as:

$$COM(\mathbf{x}) = (\ COM(\mathbf{x_{.j}}, \mathbf{x_{.t}})\ )\ j, t = 1, ..., p \tag{26}$$

The comedian matrix as a robust alternative to the covariance matrix is in general not positive (semi-) definite [Falk, 1997]. Although this problem can be fixed using Shrinkage, because it has the advantage that it will be always a positive definite and well-conditioned matrix, we adjusted the comedian matrix with the factor 2.198, resulting from the relation described in (24), which will allows us to ensure that it gives a good estimation of the covariance matrix.

Therefore, if a Shrinkage estimator is considered in (6) for the Dispersion parameter:

$$\hat{\Sigma}_{Sh} = (1-\alpha)\hat{E} + \alpha T, \tag{27}$$

we propose to use in (27), the estimator:

$$\hat{E} = \hat{S}_{CCM} = 2.198 * COM(\mathbf{x}) \tag{28}$$

In reference to the previous sub-section (15), in which we needed to provide a good estimate for $\sigma^2_{\mathbf{x_{\cdot j}}}$, for each $j = 1, ..., p$ note that, because of the relation in (24):

$$trace(\hat{S}_{CCM}) = \sum_{j=1}^{p} 2.198{*}COM(\mathbf{x_{\cdot j}}, \mathbf{x_{\cdot j}}) = \sum_{j=1}^{p} 2.198{*}MAD^2(\mathbf{x_{\cdot j}}) = \sum_{j=1}^{p} \sigma^2_{\mathbf{x_{\cdot j}}} \tag{29}$$

Thus, when considering a shrinkage estimator of the component-wise median, in order to estimate the variance of $\hat{\mu}_{CCM}$ needed in the expression (12) for the shrinkage intensity α, and according to the relation (15), we propose to estimate $\sum_{j=1}^{p} \sigma^2_{\mathbf{x_{\cdot j}}}$ using the previous relation (29).

About the shrinkage target T, several choices have been proposed in the literature. For example, [Ledoit and Wolf, 2003b] proposed a weighted average of the sample covariance matrix and a single-index covariance matrix. [Ledoit and Wolf, 2003a] proposed selecting the shrinkage target as a “constant correlation matrix”, whose correlations are set equal to the average of all sample correlations. Finally, [Ledoit and Wolf, 2004] proposed to use a multiple of the identity matrix as the shrinkage target. The authors proved that the resulting shrinkage covariance matrix is well-conditioned, even if the sample covariance matrix is not. There is also another approach introduced by [DeMiguel et al., 2013]. The authors proposed a shrinkage estimator both for the covariance matrix and its inverse. The estimators were constructed as a convex combination of the sample covariance matrix or its inverse, respectively, and a scaled shrinkage target, which they consider the identity matrix as in [Ledoit and Wolf, 2004]. Therefore, we propose to use a shrinkage target $T = \nu_\Sigma I$ as in [Ledoit and Wolf, 2004] and [DeMiguel et al., 2013], i.e. the scaled identity matrix.

Thus (27) results in:

$$\hat{\Sigma}_{R_1} = \hat{\Sigma}_{Sh(CCM)} = (1-\alpha)\hat{S}_{CCM} + \alpha\nu_\Sigma I \tag{30}$$

Lastingly, the scaling parameter ν_Σ and the shrinkage intensity parameter α in (30) needs to be estimated. They both are chosen to minimize the bias of the shrinkage target as in [Ledoit and Wolf, 2004]:

$$\begin{array}{ll} \min_{\nu_\Sigma,\alpha} & E\left[\left\|\hat{\Sigma}_{Sh} - \Sigma\right\|^2\right] \\ \text{s.t.} & \hat{\Sigma}_{Sh} = (1-\alpha)\hat{S}_{CCM} + \alpha\nu_\Sigma I, \end{array} \tag{31}$$

where $\|A\|^2 = trace(AA^t)/p$.

Proposition 3 *The scaling parameter ν_Σ and the shrinkage intensity parameter α that minimizes the expected quadratic loss, given in (31), are:*

$$\nu_\Sigma = trace(\hat{S}_{CCM})/p$$
$$\alpha = \frac{E\left[\left\|\hat{S}_{CCM} - \Sigma\right\|^2\right]}{E\left[\left\|\hat{S}_{CCM} - \nu_\Sigma I\right\|^2\right]}$$

The proof can be found in Appendix A.

In practice, we propose to estimate the nominator of the expression for α as in [Ledoit and Wolf, 2003a], [Ledoit and Wolf, 2003b] and [Ledoit and Wolf, 2004], but considering $\hat{S}_{CCM}$ instead of the sample covariance matrix, as the estimator of Σ.

A special case of comedian matrix can be defined if the data are centered using a different Location parameter. This means that, besides the option of centering at the component-wise median $\boldsymbol{\mu}_{CCM}$, which is the classical definition for the comedian matrix, we can center the data using the other Location estimators we described in (8), i.e. the multivariate L_1-median $\boldsymbol{\mu}_{MM}$, and also centering the data using the respective shrinkage estimators for both of them, which are $\boldsymbol{\mu}_{Sh(CCM)}$ and $\boldsymbol{\mu}_{Sh(MM)}$.

This also means that we can define as the Dispersion parameter the Shrinkage of those special cases of comedian matrices.

First, consider a Shrinkage over $\hat{S}_{MM} = 2.198 * COM_{MM}(\mathbf{x}) = 2.198 * (med((\mathbf{x}_{\cdot\mathbf{j}} - (\hat{\boldsymbol{\mu}}_{MM})_j)(\mathbf{x}_{\cdot\mathbf{t}} - (\hat{\boldsymbol{\mu}}_{MM})_t))$, with $j, t = 1, ..., p$. The Shrinkage estimator is then:

$$\hat{\Sigma}_{R_2} = \hat{\Sigma}_{Sh(MM)} = (1-\alpha)\hat{S}_{MM} + \alpha\nu_{\Sigma} I \tag{32}$$

Also, we can consider a Shrinkage over $\hat{S}_{Sh(CCM)} = 2.198 * COM_{Sh(CCM)}(\mathbf{x}) = 2.198 * (med((\mathbf{x}_{\cdot\mathbf{j}} - (\hat{\boldsymbol{\mu}}_{Sh(CCM)})_j)(\mathbf{x}_{\cdot\mathbf{t}} - (\hat{\boldsymbol{\mu}}_{Sh(CCM)})_t)))$, with $j, t = 1, ..., p$. Thus, a Shrinkage estimator can be defined as:

$$\hat{\Sigma}_{R_3} = \hat{\Sigma}_{Sh(Sh(CCM))} = (1-\alpha)\hat{S}_{Sh(CCM)} + \alpha\nu_{\Sigma} I \tag{33}$$

Finally, we can consider a Shrinkage over $\hat{S}_{Sh(MM)} = 2.198 * COM_{Sh(MM)}(X) = 2.198 * COM_{Sh(MM)}(\mathbf{x}) = (med((\mathbf{x}_{\cdot\mathbf{j}} - (\hat{\boldsymbol{\mu}}_{Sh(MM)})_j)(\mathbf{x}_{\cdot\mathbf{t}} - (\hat{\boldsymbol{\mu}}_{Sh(MM)})_t))$, with $j, t = 1, ..., p$. The Shrinkage estimator is then:

$$\hat{\Sigma}_{R_4} = \hat{\Sigma}_{Sh(Sh((MM))} = (1-\alpha)\hat{S}_{Sh(MM)} + \alpha\nu_{\Sigma} I \tag{34}$$

The optimal expression for the parameters α and ν_{Σ} in the above cases is analogous to the Proposition 3, but considering in each case the sample estimator as the corresponding special comedian matrix.

3.3 Combinations between location and covariance matrix estimators

A robust Mahalanobis distance can be defined as in (2), for each of the 6 possible combinations considered for the location and the covariance matrix estimators described above:

RMDv1:

$$\begin{aligned}
\widehat{\boldsymbol{\mu}}_R &= \widehat{\boldsymbol{\mu}}_{R_1} = \widehat{\boldsymbol{\mu}}_{CCM} \\
\widehat{\Sigma}_R &= \widehat{\Sigma}_{R_1} = \widehat{\Sigma}_{Sh(CCM)} = (1-\alpha)\hat{S}_{CCM} + \alpha\nu_{\Sigma} I \\
\hat{S}_{CCM} &= 2.198 * COM(\mathbf{x}) \\
&= 2.198 * (med((\mathbf{x}_{\cdot\mathbf{j}} - (\widehat{\boldsymbol{\mu}}_{R_1})_j)(\mathbf{x}_{\cdot\mathbf{t}} - (\widehat{\boldsymbol{\mu}}_{R_1})_t)))
\end{aligned}$$

RMDv2:

$$\begin{aligned}
\widehat{\boldsymbol{\mu}}_R &= \widehat{\boldsymbol{\mu}}_{R_3} = \widehat{\boldsymbol{\mu}}_{Sh(CCM)} = (1-\alpha)\hat{\boldsymbol{\mu}}_{CCM} + \alpha\nu_{\boldsymbol{\mu}}\mathbf{e} \\
\widehat{\Sigma}_R &= \widehat{\Sigma}_{R_1}
\end{aligned}$$

RMDv3:

$$\begin{aligned}
\widehat{\boldsymbol{\mu}} &= \widehat{\boldsymbol{\mu}}_{R_3} \\
\widehat{\Sigma}_R &= \widehat{\Sigma}_{R_3} = \widehat{\Sigma}_{Sh(Sh(CCM))} = (1-\alpha)\hat{S}_{Sh(CCM)} + \alpha\nu_\Sigma I \\
\hat{S}_{Sh(CCM)} &= 2.198 * COM_{Sh(CCM)}(\mathbf{x}) \\
&= 2.198 * (med((\mathbf{x}_{\cdot\mathbf{j}} - (\widehat{\boldsymbol{\mu}}_{R_3})_j)(\mathbf{x}_{\cdot\mathbf{t}} - (\widehat{\boldsymbol{\mu}}_{R_3})_t)))
\end{aligned}$$

RMDv4:

$$\begin{aligned}
\widehat{\boldsymbol{\mu}}_R &= \widehat{\boldsymbol{\mu}}_{R_2} = \widehat{\boldsymbol{\mu}}_{MM} \\
\widehat{\Sigma}_R &= \widehat{\Sigma}_{R_2} = \widehat{\Sigma}_{Sh(MM)} = (1-\alpha)\hat{S}_{MM} + \alpha\nu_\Sigma I \\
\hat{S}_{MM} &= 2.198 * COM(\mathbf{x}) \\
&= 2.198 * (med((\mathbf{x}_{\cdot\mathbf{j}} - (\widehat{\boldsymbol{\mu}}_{R_2})_j)(\mathbf{x}_{\cdot\mathbf{t}} - (\widehat{\boldsymbol{\mu}}_{R_2})_t)))
\end{aligned}$$

RMDv5:

$$\begin{aligned}
\widehat{\boldsymbol{\mu}}_R &= \widehat{\boldsymbol{\mu}}_{R_4} = \widehat{\boldsymbol{\mu}}_{Sh(MM)} = (1-\alpha)\hat{\boldsymbol{\mu}}_{MM} + \alpha\nu_{\boldsymbol{\mu}}\mathbf{e} \\
\widehat{\Sigma}_R &= \widehat{\Sigma}_{R_2}
\end{aligned}$$

RMDv6:

$$\begin{aligned}
\widehat{\boldsymbol{\mu}}_R &= \widehat{\boldsymbol{\mu}}_{R_4} \\
\widehat{\Sigma}_R &= \widehat{\Sigma}_{R_4} = \widehat{\Sigma}_{Sh(Sh(MM))} = (1-\alpha)\hat{S}_{Sh(MM)} + \alpha\nu_\Sigma I \\
S_{Sh(MM)} &= 2.198 * COM_{Sh(MM)}(\mathbf{x}) \\
&= 2.198 * (med((\mathbf{x}_{\cdot\mathbf{j}} - (\widehat{\boldsymbol{\mu}}_{R_4})_j)(\mathbf{x}_{\cdot\mathbf{t}} - (\widehat{\boldsymbol{\mu}}_{R_4})_t)))
\end{aligned}$$

For each of the combinations $RMDv5$ and $RMDv6$ described above, two versions are proposed for estimating the covariance matrix of the multivariate median $\boldsymbol{\mu}_{MM}$, needed in (21). The versions $RMDv5$ and $RMDv6$ are actually using Bootstrap for resampling the data with replacement in order to obtain 100 random samples of the same size as the original data, for each of which we calculated the multivariate median $\boldsymbol{\mu}_{MM}$ and obtained a sample for the estimator, with which we could calculate the trace of the covariance matrix. The other pair of version, $RMDv52$ and $RMDv62$, are the combinations $RMDv5$ and $RMDv6$ but using the asymptotic distribution (20), in order to estimate the covariance matrix of the multivariate median $\boldsymbol{\mu}_{MM}$.

4 Simulation results

4.1 Normal distribution

A simulation study is performed considering a $p-$dimensional random variable X following a contaminated multivariate normal distribution given as a mixture of normals of the form $(1-\alpha)N(0,I)+\alpha N(\delta\mathbf{e},\lambda I)$, where $\mathbf{e}$ denotes the $p-$dimensional vector of ones. This model is analogous to the one used by Rousseeuw and Driessen [1999], Peña and Prieto [2001], Filzmoser et al. [2005] and Peña and Prieto [2007]. This experiment has been conducted for different values of the sample-space dimension $p=2,10,30$, and the chosen sample size in relation to the dimension was $n=100,100,500$, respectively. The contamination levels were $\alpha=0,0.1,0.2,0.3,0.4$, the distance of the outliers $\delta=3,10$ and the concentration of the contamination $\lambda=0.1,1$. For each set of values, 100 random sample repetitions have been generated.

For the methods mentioned in previous sections some measures are studied: the correct classification rates (c) and the false classification rates (f). These measures are shown in the Tables 4-15, for each contamination scheme. In the tables, the method *Classic* refers to the classical Mahalanobis distance (1), the method *MCD* refers to the robust Mahalanobis distance based on the MCD estimators (3), the method *Adj.MCD* refers to the latter distance considering the adjusted quantile of Filzmoser et al. [2005], the method *Kurtosis* refers to the Peña and Prieto [2007] approach which makes use of the distance defined in (5). The other eight methods are the collection proposed in this paper and described in the previous section.

The results from the tables show general outcomes. *Classic* is the method with the worst performance with respect to "c" (correct detection rate) in all cases. It can also be seen that for the Mahalanobis distance with the *MCD*, the use of the adjusted quantile as the threshold, actually improves the method with respect to the "f" (false detection rate), lowering it. However, although in some cases it maintains the same "c", in other cases it also lowers the "c".

In the case of dimension p = 2 (Tables 4-7), when there is no contamination (i.e. $\alpha=0$) all methods show a sufficiently low "f" value. Regarding our collection of methods, when the outliers are closer to the central data (i.e. $\delta=3$), this results in a value "c" somewhat lower than that of our competitors. On the other hand, when the outliers are farther ($\delta=10$), our methods show a "c" of 100% in the vast majority of cases, and an "f" sufficiently low and close to zero.

In the case of dimension p = 10 (Tables 8-11), our methods show the lowest values of “f”, without taking into account for *ClassicMaha*, which is the method with worse results, in fact unacceptable results in all cases of contamination from 10% to 40%. When atypical values are closer to the central data ($\delta = 3$), our methods have the best performance for 10% and 20% levels of contamination, and for the remaining two levels, 30% and 40%, they show a lower “c” than that of our competitors, although on the other hand, they also show a lower “f”. In the case of more distant outliers ($\delta = 10$), our methods present the best behavior in all cases, and in the vast majority of cases with a “c” of 100% and “f” equal to zero, or close enough to zero.

In the case of dimension p = 30 (Tables 12-15), the behavior of our methods is very similar to the previous case of dimension $p = 10$, but it should be noted that the performance of our competitors is much worse, mainly from the 10% or 20% of contamination to the 40%. Only the *Kurtosis* method has an exception in two cases in which it shows a higher “c”, but also a higher “f”, in the case of nearest outliers ($\delta = 3$) and with 40% of contamination.

The results from the tables can be seen graphically for each contamination scheme in order to see the performance for each method through the increasing contamination. The Figures 1-4 show the results when the dimension $p = 2$. The Figures 5-8 show the results when the dimension $p = 10$. The Figures 9-12 show the results when the dimension $p = 30$. The desirable behavior would be the higher possible value for the correct classification rates (c) and the lowest possible value for the false classification rates (f). It should be noted that in the figures, the considered methods are our competitors and the last five combinations from the proposed collection, which are $RMDv4$, $RMDv5$, $RMDv52$, $RMDv6$, $RMDv62$. Note that, versions $RMDv52$ and $RMDv62$ are actually versions 5 and 6 considering the asymptotic distribution of the multivariate median, instead of considering Bootstrap. The latter decision is motivated for the results in the tables and because these versions are the ones that considered the multivariate L_1 median, which captures better the multivariate characteristic of the data.

4.1.1 Computational times

The results from the Tables 16-18 with the computational times show that the fastest method in all cases is the *Classic Mahalanobis*, which is expected because it is the simplest method. It can also be seen that the versions *RMDv5* and *RMDv6* corresponding to using Bootstrap are our slower combinations. However for these two combinations using the asymptotic distribution of the multivariate median, which are *RMDv52* and *RMDv62*, and also for the combination *RMDv4*, the results are competitive. Compared to the *MCD*, the latter is between 6 and 21 times slower than our three methods (*RMDv4*, *RMDv52* and *RMDv62*). The *MCD* version with the adjusted quantile has similar behavior to its unadjusted version. And finally, the *Kurtosis* method has computational times similar to ours in dimension $p = 2$ and $p = 10$, while in dimension $p = 30$ it is 3 or 4 times slower.

4.2 T_3 distribution

In order to check the behavior of the methods when the distribution deviates from normality, a simulation study is performed considering a $p-$dimensional random variable X following a contaminated multivariate T-distribution with 3 degrees of freedom of the form $(1-\alpha)T_3(0, I) + \alpha T_3(\delta\mathbf{e}, \lambda I)$, where $\mathbf{e}$ denotes the $p-$dimensional vector of ones. The first parameter of the notation of $T_3(\cdot, \cdot)$ refers to the mean and the second one to the covariance matrix. The parameters for the contamination are the same considered above.

For the methods mentioned in previous sections, the sames measures as in the previous section are studied: the correct classification rates (c) and the false classification rates (f). The results can be found in the Tables 19-30. The tables show that *Classic Mahalanobis* is again the method with worst performance throughout all the schemes of contamination.

For the dimension $p = 2$ (Tables 19- 22), when the outliers are near the center of the data ($\delta = 3$) and are concentrated ($\lambda = 0.1$), the *Kurtosis* method is the one with the highest "c" when the level of contamination increases, although all methods show the same behavior, i.e. they all decrease their "c" throughout the contamination. On the other hand, *Kurtosis* is also the method with highest "f", and all methods increase their "f" throughout the contamination, except for our methods, whose "f" decreases. When we consider near but less concentrated outliers ($\lambda = 1$), our methods have a "c" value better than the other methods for a 10% and a 20% level of contamination. However, for a 30% and 40% of contamination, *Kurtosis* presents a higher "c", but our methods present smaller "f". For more distant outliers ($\delta = 10$), and up to 30% of contamination, the correct classification rate "c" of our methods remains at 1, or very close to 1, while the false classification rate "f" decreases along the contamination. For a 40% level of contamination, *Kurtosis* has a better "c" but also a worst "f".

For dimension $p = 10$ (Tables 23- 26), and for outliers near the center, our methods show the highest value of "c" and the smallest values of "f" in most cases. For example, in the case of less concentrated outliers ($\lambda = 1$), and a 40% of contamination, *Kurtosis* shows the highest "c", although it is not a very desirable value, apart from the fact of having a fairly high "f" too. For more distant outliers, our methods behave even better, with "c" rates equal to 1 except for the 40% of contamination, where they show a "c" very similar to the *Kurtosis* method. And on the other hand, the "f" decreases through the contamination.

In dimension $p = 30$ (Tables 27- 30), considering the case of atypical values near the center and also concentrated, the behavior of our methods is the most favorable in comparison with the others, although from a 20% of contamination, the "c" decreases considerably. However, the behavior of our competitors is unfavorable in all cases, with respect to both "c" and "f". When the outliers are close but not as much concentrated, our methods have the highest values of "c" from 10% to 30% of contamination, and the "f" values decrease. For *Kurtosis*, only at 40% of contamination it has a better "c", but also a worse "f". For the case of farther contamination, our methods have values of "c" equal to 1, it only decreases very little when there is a 40% of contamination, while the values of "f" decrease. On the other hand, our competitors have very unfavorable behaviors, with respect to "c" and "f".

It should be noted that in these schemes we can observe the unsatisfactory behavior of our competitors, and that in most cases, our methods have the best "c" and "f" values. In fact, on a few occasions we do not have the best "c", we do have the best "f". And we must bear in mind that in the cases we lose about the "c", we do it against levels "c" of the competitors which are not very desirable either.

The correct classification rates (c) and the false classification rates (f), can be seen graphically for each contamination scheme in order to see the performance for each method through the increasing contamination, as in the previous section. Tables 13-16 show the results for dimension $p = 2$. Tables 17-20 show the results for dimension $p = 10$. Tables 21-24 show the results for dimension $p = 30$.

4.3 Exponential distribution

A simulation study is performed considering a $p-$dimensional random variable X following a contaminated multivariate Exponential distribution given as a mixture $(1-\alpha)Exp(0) + \alpha Exp(\delta \mathbf{e})$, where $\mathbf{e}$ denotes the $p-$dimensional vector of ones. The parameter of the notation $Exp(\cdot)$ refers to the mean. This case is analogous to the previous ones, with the difference that only the schemes associated with the distance of the outliers are considered. This means that the parameters p, n, α, δ

and the number of repetitions remains the same as above, but the concentration of the contamination do not varies.

For the methods mentioned in previous sections the same measures are studied: the correct classification rates (c) and the false classification rates (f). These measures are shown in the Tables 31-36. It can be seen that *Classic Mahalanobis* is again the one with worse performance throughout all the schemes of contamination.

For $p = 2$ (Tables 31-32), and near contamination, it is difficult to choose the method with the best behavior, because the method with higher "c" also has higher "f", which is *Kurtosis*. While the one that presents smaller "f" also presents smaller "c", which is the *Adjusted MCD*. Our methods are in 2nd place for the best "c" and "f", but they are the best if we are looking for a balance.

For $p = 10$ (Tables 33-34), our methods show the highest values for the "c", with values equal to 1, or very close to 1. In these cases, *Kurtosis* has similar "c" values but slightly below, and his "f" is worse than our methods. For both versions of the *MCD*, their "f" is the best but their "c" is the worst, and quite unfavorable compared to our methods.

For $p = 30$ (Tables 35-36), our methods have the best value for "c" , while "f" decreases, although they are not the best regarding the latter. The *Adjusted MCD* has the best "f" rate, although it also has the worst "c".

The correct classification rates (c) and the false classification rates (f), can be seen graphically for each contamination scheme in order to see the performance for each method through the increasing contamination, as in the previous section. Tables 25-26 show the results when the dimension is $p = 2$. Tables 27-28 show the results when the dimension is $p = 10$. Tables 29-30 show the results when the dimension is $p = 30$.

4.4 Summary about the Simulations

In the simulation study, for each contamination scheme we have also calculated a measure called F-score (Goutte and Gaussier [2005], Sokolova et al. [2006], Powers [2011]), often used in Engineering, which is a measure of a test's accuracy. Its expression is F-score= $2PR/(P + R)$, where P is called precision and R is known as the recall. The precision P is the number of correct detected outliers divided by the total number of detected outliers, and the recall R is the number of correct detected outliers divided by the real total number of outliers. Thus, this measure provides a trade-off between the two desired outcomes: a high rate of correctly identified outliers and a low rate of observations mislabel as outliers. The results are not included in the paper for avoiding large extension, but in summary the method with the overall classification between the top 3 best positions ranking with respect to the F-score, is method $RMDv62$, a result that can also be seen in the previous simulation study. Another interesting aspect to mention here is the good performance of our proposed methods when we deviate from the normality assumption, for example when considering skewed and heavy-tailed distributions like the multivariate T distribution and the multivariate Exponential distribution.

5 Real dataset

The collection of proposed methods are applied to a real dataset to evaluate their performance. The following dataset was taken from the *UCI Knowledge Discovery in Databases Archive* [Bay, 1999]. Specifically, we have chosen the *Breast Cancer Wisconsin (Diagnostic) Data Set* (WDBC). Features are computed from a digitized image of a fine needle aspirate of a breast mass. They describe 30 characteristics of the cell nuclei present in the image, for 569 samples, from which 357 are benign and 212 malign. We propose to study only the 357 benign data, as in [Maronna and Zamar, 2002]. Therefore, this example has dimension $p = 30$ and sample size $n = 357$. We applied each method for detecting outliers and we retained the results, along with the computational times. In order to interpret the outcome, we show the normalized data (we have standardized the results after the detection only for better visualization aim). We have also plotted the multivariate L_1 median and a kind of multivariate "boxplot", which is an adaptation from Sun and Genton [2011] method, but for finite dimensional. What the "box" would be is constructed sorting the data according to their L_1 depth value. The corresponding Q_1 and Q_3 "quartiles" delimiting the "box" are in fact the minimum and maximum values for each coordinate taking only into account the 50% of the most central data. Thus, the "fences" can be constructed with the same approach $F1 = Q_1 - 1.5RI$ and $F2 = Q_3 + 1.5RI$, where the "interquartile range" is $RI = Q_3 - Q_1$. Then, we can look for each method's result how many detected outliers are inside the "fences" for all their coordinates, and how many are outside the "fences".

The Figures 31-37 show the results graphically, for each method, in parallel coordinates (Inselberg and Dimsdale [1990], Wegman [1990], Inselberg [2009]). In blue color is the not outlying data, in yellow color is the "box" delimiting the 50% of most central data, in red is the "fences" and in "cyan" is the multivariate median. Also, in green color are the outliers inside the "fences" and in magenta the outliers outside the "fences". We consider from our collection, only the version *RMDv4* and the versions *RMDv52* and *RMDv62* which are combinations 5 and 6 without considering Bootstrap. There are just a few outside the "fences", which are clearly outliers. In the following table we show the results for each method.

Table 1: Detected outliers inside and outside the fences.

Method	**Out_In**	**Out_out**	**Out_total**
ClassicMaha	37	4	41
MCD	72	4	76
AdjustedMCD	64	4	68
Kurtosis	155	4	159
RMDv4	31	3	34
RMDv52	28	4	32
RMDv62	25	3	28

In the Table 1, it can be seen that outside the "fences" there are 3 or 4 for all the methods. Also, the method *Kurtosis* detected 159 outliers out of the 357 data. Furthermore, our three methods are the ones that detect less amount of data as outliers.

The following table shows the result to see how many of the detected outliers belong to the 50% of the most central data, according to the L_1-median.

Table 2: Detected outliers inside and outside the 50% of the most central data.

Method	**Out_In**	**Out_out**	**Out_total**
ClassicMaha	11	30	41
MCD	29	47	76
AdjustedMCD	27	41	68
Kurtosis	63	96	159
RMDv4	9	25	34
RMDv52	8	24	32
RMDv62	7	21	28

As it can be seen the Table 2, our three methods are the ones that detect less amount of the 50% of the most central data as outliers. We can investigate the shape of the detected outliers that belong to the 50% of the most central data, in order to see if they are similar or near to the median, or if they have a distinct shape and in fact they could be outliers.

The figures 38-48 show the shape for the detected outliers that belong to the 50% of the most central data (or a selection of them in case of too many). In cyan color is the multivariate median, in yellow color the "box" and in blue color the detected outlier.

Through the analysis of the figures corresponding to the atypical values detected by our competitors, it can be seen that for all those methods there seem to be some outliers that have a shape very similar to the multivariate median or they are close to it for all the values of its components, leading us to think that maybe they are detecting too much. However, in the figures associated with our three methods, we can see that all outliers are quite different than the multivariate median, in fact, they might be shape outliers. For a final argument, we can say that all the outliers that belong to the 50% of the most central data detected by the method we consider the best, *RMDv62*, are actually detected by our competitors (the intersection matches), so this makes us think that our method detects just enough.

Table 3: Computational times in seconds for each methods with the WDBC dataset.

Methods	**Computational Time (in sec)**
ClassicMaha	0,087201656
MCD Maha	8,907105534
Adjusted MCD	8,256962981
Kurtosis	6,166659664
RMD SH3 v4	1,350939047
RMD SH3 v5	92,23925777
RMD SH3 v52	1,198164309
RMD SH3 v6	92,3412066
RMD SH3 v62	1,132937594

Table 3 show the computational times for each method in the task of detecting outliers with this example of real dataset. The results demonstrates how in this example, without considering *ClassicMaha* (the fastest method), the rest of our competitors are between 6 or 8 times slower than our three methods *RMDv4*, *RMDv52* and *RMDv62*, approximately.

6 Conclusions

Correct detection of outliers in the multivariate case is well-known to be a very important task for thorough data analysis. In order to reach that goal properly, it is necessary to consider the shape of the data and its structure in the multivariate space. That is the reason why the Mahalanobis distance approach is frequently used for the task of identifying the outliers. Different robust Mahalanobis distances can be defined according to the selected location and dispersion parameters. There are various well-known definitions in the literature that we have considered in this paper: the classical Mahalanobis distance, the MCD based Mahalanobis distance, the latter using the adjusted quantile as the threshold for detecting outliers, and the Kurtosis procedure. We have proposed a collection of different combinations of robust location and covariance matrix estimators based on the notion of Shrinkage, in order to define with each combination a robust Mahalanobis distance. The performance for the proposed collection and the other outlier detection methods from the literature is shown with a simulation study. It can be concluded that the classical Mahalanobis distance is clearly not effective in identifying multiple outliers. Also, the other approaches from the literature, which consider robust distances, do not have a very desirable behavior through all the schemes of contamination, specially in high dimension. While the proposed collection of robust distances have the ability to discover outliers with high precision in the vast majority of cases in the simulations. The results, specially for high dimension, about the correct and false classification rates, the behavior with skewed or heavy-tailed distributions and our inexpensive computational times shows the competitiveness of our proposal. A real dataset example is also studied, in which the results bear out the latter conclusions.

The results presented in this article emphasize the advantages of using Shrinkage estimators for the Location and Covariance matrix parameters in the definition of a robust Mahalanobis distance. It remains to be examined whether the proposal could be improved by adapting the adjusted quantile to the proposed robust distances. It could also be an interesting matter to study, whether the use of the different definitions of "depth" in the literature (Tukey [1975], Liu et al. [1990], Serfling [2002], Chen et al. [2009], Agostinelli and Romanazzi [2011], Paindaveine and Van Bever [2013]), could improve the performance of the approach, as it is known that depth is a robust measure for location.

7 Acknowledgment

This research was partially supported by Spanish Ministry grant ECO2015-66593-P.

Appendix A Proofs

Proof of Proposition 1.

The optimization problem is:

$$\begin{array}{cc} \min_{\nu_{\boldsymbol{\mu}},\alpha} & E\left[\left\|\boldsymbol{\mu}_{Sh(CCM)}-\boldsymbol{\mu}\right\|_2^2\right] \\ \text{s.t.} & \boldsymbol{\mu}_{Sh(CCM)}=(1-\alpha)\boldsymbol{\mu}_{CCM}+\alpha\nu_{\boldsymbol{\mu}}\mathbf{e}, \end{array} \tag{35}$$

where $\|\mathbf{x}\|_2^2=\sum_{j=1}^p x_j^2$ and the associated inner product is: $\langle x,y\rangle=\sum_{j=1}^p x_j y_j$.

The objective function is equivalent to:

$$\begin{aligned} E\left[\left\|\boldsymbol{\mu}_{Sh(CCM)}-\boldsymbol{\mu}\right\|_2^2\right] &= E\left[\|(1-\alpha)\boldsymbol{\mu}_{CCM}+\alpha\nu_{\boldsymbol{\mu}}\mathbf{e}-\boldsymbol{\mu}\|_2^2\right] \\ &= (1-\alpha)^2 E\left[\|\boldsymbol{\mu}_{CCM}-\boldsymbol{\mu}\|_2^2\right]+\alpha^2\left\|\nu_{\boldsymbol{\mu}}\mathbf{e}-\boldsymbol{\mu}\right\|_2^2 \\ &+ 2E\left[\langle(1-\alpha)(\boldsymbol{\mu}_{CCM}-\boldsymbol{\mu}),\alpha(\nu_{\boldsymbol{\mu}}\mathbf{e}-\boldsymbol{\mu})\rangle\right] \end{aligned}$$

The latter element in the above expression is equal to zero because $E(\boldsymbol{\mu}_{CCM})=\boldsymbol{\mu}$ (see [Chu, 1955]). Then, the optimization problem (35) reduces to minimize:

$$E\left[\left\|\boldsymbol{\mu}_{Sh(CCM)}-\boldsymbol{\mu}\right\|_2^2\right]=(1-\alpha)^2 E\left[\|\boldsymbol{\mu}_{CCM}-\boldsymbol{\mu}\|_2^2\right]+\alpha^2\left\|\nu_{\boldsymbol{\mu}}\mathbf{e}-\boldsymbol{\mu}\right\|_2^2 \tag{36}$$

In order to find the optimal $\nu_{\boldsymbol{\mu}}$, it is necessary to minimize only the right element of the above expression.

$$||\nu_{\boldsymbol{\mu}}\mathbf{e}-\boldsymbol{\mu}||_2^2=\nu_{\boldsymbol{\mu}}^2\left\|\mathbf{e}\right\|_2^2+\left\|\boldsymbol{\mu}\right\|_2^2-2\nu_{\boldsymbol{\mu}}\left\langle\mathbf{e},\boldsymbol{\mu}\right\rangle \tag{37}$$

Then, with respect to the scaling parameter, the first order optimality condition give:

$$0=2p\nu_{\boldsymbol{\mu}}-2\left\langle\mathbf{e},\boldsymbol{\mu}\right\rangle=2\left(p\nu_{\boldsymbol{\mu}}-\sum_{j=1}^p\boldsymbol{\mu}_j\right) \tag{38}$$

Thus:

$$\nu_{\boldsymbol{\mu}}=\frac{1}{p}\sum_{j=1}^p\boldsymbol{\mu}_j \tag{39}$$

Estimating $\boldsymbol{\mu}$ as $\boldsymbol{\mu}_{CCM}$, we obtain:

$$\nu_{\boldsymbol{\mu}} = \frac{\hat{\boldsymbol{\mu}}_{CCM}\mathbf{e}}{p}$$

In (36), with respect to the shrinkage intensity parameter α, the first order optimality condition give:

$$0 = 2(1-\alpha)E\left[\|\boldsymbol{\mu}_{CCM} - \boldsymbol{\mu}\|_2^2\right] + 2\alpha\left\|\nu_{\boldsymbol{\mu}}\mathbf{e} - \boldsymbol{\mu}\right\|_2^2$$

Hence:

$$\alpha = \frac{E\left[\|\hat{\boldsymbol{\mu}}_{CCM} - \boldsymbol{\mu}\|_2^2\right]}{E\left[\|\hat{\boldsymbol{\mu}}_{CCM} - \nu_{\boldsymbol{\mu}}\mathbf{e}\|_2^2\right]} \tag{40}$$

Proof of Proposition 2.

The optimization problem is:

$$\begin{array}{cc} \min_{\nu_{\boldsymbol{\mu}},\alpha} & E\left[\left\|\boldsymbol{\mu}_{Sh(MM)} - \boldsymbol{\mu}\right\|_2^2\right] \\ \text{s.t.} & \boldsymbol{\mu}_{Sh(MM)} = (1-\alpha)\boldsymbol{\mu}_{MM} + \alpha\nu_{\boldsymbol{\mu}}\mathbf{e}, \end{array} \tag{41}$$

where $\|x\|_2^2 = \sum_{j=1}^p x_j^2$.

Similarly to the previous demonstration, we can consider the following expression for the objective function:

$$\begin{aligned} E\left[\left\|\boldsymbol{\mu}_{Sh(MM)} - \boldsymbol{\mu}\right\|_2^2\right] &= E\left[\|(1-\alpha)\boldsymbol{\mu}_{MM} + \alpha\nu_{\boldsymbol{\mu}}\mathbf{e} - \boldsymbol{\mu}\|_2^2\right] \\ &= (1-\alpha)^2 E\left[\|\boldsymbol{\mu}_{MM} - \boldsymbol{\mu}\|_2^2\right] + \alpha^2\left\|\nu_{\boldsymbol{\mu}}\mathbf{e} - \boldsymbol{\mu}\right\|_2^2 \\ &+ 2E\left[\langle(1-\alpha)(\boldsymbol{\mu}_{MM} - \boldsymbol{\mu}), \alpha(\nu_{\boldsymbol{\mu}}\mathbf{e} - \boldsymbol{\mu})\rangle\right] \end{aligned}$$

The expectation of the inner product is equal to zero because in [Bose and Chaudhuri, 1993] and [Bose, 1995] the authors investigated the asymptotic distribution for the L_1-median, and they obtained the following result about the covariance matrix in presence of normality:

$$\boldsymbol{\mu}_{MM} \sim N_p\left(\boldsymbol{\mu}, \frac{1}{n}\hat{A}^{-1}\hat{B}\hat{A}^{-1}\right), \tag{42}$$

Then, the optimization problem (41) reduces to minimize:

$$E\left[\left\|\boldsymbol{\mu}_{Sh(MM)}-\boldsymbol{\mu}\right\|_2^2\right]=(1-\alpha)^2E\left[\|\boldsymbol{\mu}_{MM}-\boldsymbol{\mu}\|_2^2\right]+\alpha^2\left\|\nu_{\boldsymbol{\mu}}\mathbf{e}-\boldsymbol{\mu}\right\|_2^2 \tag{43}$$

Then, the optimal parameter $\nu_{\boldsymbol{\mu}}$ can be found minimizing only the right element of the above expression, which is the only one depending on that parameter.

$$||\nu_{\boldsymbol{\mu}}\mathbf{e}-\boldsymbol{\mu}||_2^2=\nu_{\boldsymbol{\mu}}^2\left\|\mathbf{e}\right\|_2^2+\left\|\boldsymbol{\mu}\right\|_2^2-2\nu_{\boldsymbol{\mu}}\left\langle\mathbf{e},\boldsymbol{\mu}\right\rangle \tag{44}$$

The associated first order optimality condition give:

$$0=2p\nu_{\boldsymbol{\mu}}-2\left\langle\mathbf{e},\boldsymbol{\mu}\right\rangle=2\left(p\nu_{\boldsymbol{\mu}}-\sum_{j=1}^{p}\boldsymbol{\mu}_j\right) \tag{45}$$

Therefore:

$$\nu_{\boldsymbol{\mu}}=\frac{1}{p}\sum_{j=1}^{p}\boldsymbol{\mu}_j \tag{46}$$

In practice, we propose to estimate $\boldsymbol{\mu}$ with $\boldsymbol{\mu}_{MM}$. Thus:

$$\nu_{\boldsymbol{\mu}}=\frac{\hat{\boldsymbol{\mu}}_{MM}\mathbf{e}}{p}$$

With respect to the shrinkage intensity parameter α, the first order optimality condition associated to (43), give:

$$0=2(1-\alpha)E\left[\|\boldsymbol{\mu}_{MM}-\boldsymbol{\mu}\|_2^2\right]+2\alpha\left\|\nu_{\boldsymbol{\mu}}\mathbf{e}-\boldsymbol{\mu}\right\|_2^2$$

Hence:

$$\alpha=\frac{E\left[\|\hat{\boldsymbol{\mu}}_{MM}-\boldsymbol{\mu}\|_2^2\right]}{E\left[\|\hat{\boldsymbol{\mu}}_{MM}-\nu_{\boldsymbol{\mu}}\mathbf{e}\|_2^2\right]} \tag{47}$$

Proof of Proposition 3.

The optimization problem is:

$$\begin{array}{cc}\min_{\nu_\Sigma,\alpha} & E\left[\|\Sigma_{Sh}-\Sigma\|^2\right] \\ \text{s.t.} & \Sigma_{Sh}=(1-\alpha)S_{CCM}+\alpha\nu_\Sigma I,\end{array} \tag{48}$$

where $\|A\|^2=trace(AA^t)/p$, and the associated inner product is $\langle A_1,A_2\rangle=trace(A_1A_2^t)/p$

Analogous to the previous Propositions, the objective function in the above minimization problem (48) can be seen as:

$$
\begin{aligned}
E\left[\|\Sigma_{Sh}-\Sigma\|^2\right] &= E\left[\|(1-\alpha)S_{CCM}+\alpha\nu_\Sigma I-\Sigma\|^2\right] \\
&= (1-\alpha)^2 E\left[\|S_{CCM}-\Sigma\|^2\right]+\alpha^2\left\|\nu_\Sigma I-\Sigma\right\|^2 \\
&+ 2E\left[\langle(1-\alpha)(S_{CCM}-\Sigma),\alpha(\nu_\Sigma I-\Sigma)\rangle\right]
\end{aligned}
$$

In this case, note that the latter element in the above expression is equal to zero because $E(\hat{S}_{CCM})=\Sigma$ (24). Hence, the optimization problem (48) reduces to minimize the following expression:

$$
E\left[\|\Sigma_{Sh}-\Sigma\|^2\right] = (1-\alpha)^2 E\left[\|S_{CCM}-\Sigma\|^2\right]+\alpha^2\left\|\nu_\Sigma I-\Sigma\right\|^2 \tag{49}
$$

The optimal ν_Σ can be obtained by minimizing only the right element of the above expression, because it is the only one depending on that parameter. Also, note that:

$$
||\nu_\Sigma I-\Sigma||^2 = \nu_\Sigma^2\left\|I\right\|^2+\left\|\Sigma\right\|^2-2\nu_\Sigma\left\langle I,\Sigma\right\rangle \tag{50}
$$

Then, the first order optimality condition with respect to the scaling parameter, give:

$$
\begin{aligned}
0 &= 2\nu_\Sigma-2\left\langle I,\Sigma\right\rangle \\
\nu_\Sigma &= \left\langle I,\Sigma\right\rangle = trace(\Sigma I^t)/p
\end{aligned}
$$

Therefore:

$$
\nu_\Sigma = trace(\Sigma)/p
$$

In practice, we propose to estimate Σ with $\hat{S}_{CCM}$, thus:

$$
\nu_\Sigma = trace(\hat{S}_{CCM})/p
$$

In (49), with respect to the shrinkage intensity parameter α, the first order optimality condition give:

$$
\alpha = \frac{E\left[\left\|\hat{S}_{CCM}-\Sigma\right\|^2\right]}{E\left[\left\|\hat{S}_{CCM}-\nu_\Sigma I\right\|^2\right]} \tag{51}
$$

Appendix B Tables

Multivariate Normal distribution

The Tables 4-15 show the correct classification rates (c) and the false classification rates (f) for each method, corresponding to the simulations with multivariate Normal distribution explained in Section 4.1. Note that for $\alpha = 0$, there is no contamination, thus the first column corresponding to the "c" rate in this case is "NaN".

Table 4: Normal distribution $p = 2$, $n = 100$, $\delta = 3$, $\lambda = 0.1$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,0212	0,2295	0,0126	0,0075	0,0189	0,0000	0,0258	0,0000	0,0423
MCD	NaN	0,0426	0,9649	0,0177	0,7416	0,0172	0,0799	0,1209	0,0003	0,1772
Adj.MCD	NaN	0,0267	0,9351	0,0072	0,6801	0,0132	0,0716	0,0978	0,0000	0,1472
Kurtosis	NaN	0,0185	0,9692	0,0195	0,7193	0,0249	0,2000	0,0433	0,0000	0,2561
RMDv1	NaN	0,0319	0,9507	0,0107	0,5090	0,0043	0,0308	0,0004	0,0000	0,0008
RMDv2	NaN	0,0311	0,9468	0,0102	0,5032	0,0042	0,0292	0,0004	0,0000	0,0002
RMDv3	NaN	0,0304	0,9539	0,0104	0,4892	0,0042	0,0344	0,0004	0,0000	0,0000
RMDv4	NaN	0,0330	0,9522	0,0116	0,5023	0,0043	0,0522	0,0005	0,0000	0,0103
RMDv5	NaN	0,0311	0,9525	0,0108	0,4912	0,0041	0,0470	0,0004	0,0000	0,0089
RMDv52	NaN	0,0312	0,9523	0,0113	0,4944	0,0042	0,0485	0,0004	0,0000	0,0091
RMDv6	NaN	0,0300	0,9420	0,0114	0,4860	0,0043	0,0433	0,0004	0,0000	0,0079
RMDv62	NaN	0,0302	0,9509	0,0116	0,4922	0,0043	0,0449	0,0004	0,0000	0,0079

Computational times from the simulation study with a multivariate Normal distribution.

Table 5: Normal distribution $p = 2$, $n = 100$, $\delta = 10$, $\lambda = 0.1$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,0228	0,7358	0,0088	0,0241	0,0130	0,0000	0,0238	0,0000	0,0350
MCD	NaN	0,0425	1,0000	0,0169	0,9200	0,0035	0,1400	0,1020	0,0000	0,1871
Adj.MCD	NaN	0,0261	1,0000	0,0056	0,9200	0,0012	0,1400	0,0839	0,0000	0,1566
Kurtosis	NaN	0,0203	1,0000	0,0222	0,8900	0,0178	0,9097	0,0234	0,9500	0,0654
RMDv1	NaN	0,0310	1,0000	0,0142	1,0000	0,0035	1,0000	0,0004	0,7986	0,0000
RMDv2	NaN	0,0312	1,0000	0,0142	1,0000	0,0037	1,0000	0,0004	0,7958	0,0000
RMDv3	NaN	0,0307	1,0000	0,0139	1,0000	0,0033	1,0000	0,0003	0,7934	0,0000
RMDv4	NaN	0,0309	1,0000	0,0148	1,0000	0,0033	1,0000	0,0004	0,8425	0,0000
RMDv5	NaN	0,0301	1,0000	0,0134	1,0000	0,0039	1,0000	0,0003	0,8340	0,0000
RMDv52	NaN	0,0302	1,0000	0,0141	1,0000	0,0038	1,0000	0,0003	0,8351	0,0000
RMDv6	NaN	0,0297	1,0000	0,0142	1,0000	0,0032	1,0000	0,0003	0,8284	0,0000
RMDv62	NaN	0,0296	1,0000	0,0137	1,0000	0,0034	1,0000	0,0003	0,8224	0,0000

Table 6: Normal distribution $p = 2$, $n = 100$, $\delta = 3$, $\lambda = 1$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,0225	0,3857	0,0091	0,1048	0,0088	0,0356	0,0080	0,0189	0,0104
MCD	NaN	0,0455	0,9007	0,0170	0,6546	0,0028	0,2056	0,0110	0,0601	0,0352
Adj.MCD	NaN	0,0290	0,8307	0,0082	0,5716	0,0010	0,1526	0,0072	0,0355	0,0240
Kurtosis	NaN	0,0206	0,9084	0,0203	0,7994	0,0237	0,3641	0,0133	0,1160	0,0100
RMDv1	NaN	0,0343	0,8305	0,0139	0,5290	0,0047	0,1729	0,0008	0,0182	0,0003
RMDv2	NaN	0,0336	0,8308	0,0135	0,5250	0,0049	0,1700	0,0008	0,0166	0,0003
RMDv3	NaN	0,0342	0,8282	0,0130	0,5278	0,0052	0,1643	0,0007	0,0147	0,0001
RMDv4	NaN	0,0318	0,8220	0,0129	0,5176	0,0047	0,1920	0,0008	0,0187	0,0003
RMDv5	NaN	0,0324	0,8217	0,0122	0,5178	0,0051	0,1875	0,0010	0,0163	0,0003
RMDv52	NaN	0,0322	0,8217	0,0125	0,5167	0,0051	0,1906	0,0008	0,0172	0,0003
RMDv6	NaN	0,0327	0,8292	0,0119	0,5217	0,0048	0,1794	0,0008	0,0148	0,0001
RMDv62	NaN	0,0329	0,8307	0,0125	0,5207	0,0049	0,1878	0,0008	0,0151	0,0001

Table 7: Normal distribution $p = 2$, $n = 100$, $\delta = 10$, $\lambda = 1$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,0225	0,7295	0,0075	0,1206	0,0074	0,0248	0,0079	0,0141	0,0092
MCD	NaN	0,0464	1,0000	0,0151	0,9200	0,0032	0,2689	0,0110	0,0584	0,0485
Adj.MCD	NaN	0,0292	1,0000	0,0068	0,9200	0,0005	0,2642	0,0070	0,0387	0,0345
Kurtosis	NaN	0,0205	1,0000	0,0223	0,9904	0,0215	0,8606	0,0204	0,8298	0,0433
RMDv1	NaN	0,0305	1,0000	0,0115	1,0000	0,0029	1,0000	0,0001	0,8297	0,0023
RMDv2	NaN	0,0303	1,0000	0,0124	1,0000	0,0030	1,0000	0,0001	0,8285	0,0023
RMDv3	NaN	0,0311	1,0000	0,0112	1,0000	0,0029	1,0000	0,0001	0,8251	0,0023
RMDv4	NaN	0,0327	1,0000	0,0116	1,0000	0,0025	1,0000	0,0004	0,8702	0,0056
RMDv5	NaN	0,0320	1,0000	0,0117	1,0000	0,0024	1,0000	0,0003	0,8629	0,0052
RMDv52	NaN	0,0321	1,0000	0,0115	1,0000	0,0025	1,0000	0,0003	0,8659	0,0052
RMDv6	NaN	0,0323	1,0000	0,0117	1,0000	0,0028	1,0000	0,0000	0,8526	0,0039
RMDv62	NaN	0,0329	1,0000	0,0120	1,0000	0,0027	1,0000	0,0003	0,8549	0,0039

Table 8: Normal distribution $p = 10$, $n = 100$, $\delta = 3$, $\lambda = 0.1$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,0197	0,0125	0,0207	0,0000	0,0405	0,0000	0,0838	0,0000	0,1588
MCD	NaN	0,1205	0,8100	0,0883	0,0400	0,1803	0,0000	0,2464	0,0000	0,3247
Adj.MCD	NaN	0,1011	0,8100	0,0676	0,0400	0,1558	0,0000	0,2194	0,0000	0,2935
Kurtosis	NaN	0,0740	0,7000	0,1443	0,3096	0,3274	0,7387	0,1672	0,5292	0,2949
RMDv1	NaN	0,0381	1,0000	0,0076	0,6107	0,0006	0,0044	0,0000	0,0000	0,0012
RMDv2	NaN	0,0390	1,0000	0,0071	0,6000	0,0008	0,0035	0,0000	0,0000	0,0008
RMDv3	NaN	0,0354	0,9994	0,0074	0,5730	0,0003	0,0035	0,0000	0,0000	0,0006
RMDv4	NaN	0,0253	0,9993	0,0045	0,7827	0,0002	0,1565	0,0000	0,0226	0,0411
RMDv5	NaN	0,0190	0,9953	0,0028	0,5744	0,0002	0,0711	0,0000	0,0007	0,0404
RMDv52	NaN	0,0239	0,9993	0,0041	0,7718	0,0002	0,1441	0,0000	0,0183	0,0408
RMDv6	NaN	0,0348	0,9813	0,0062	0,6191	0,0005	0,0872	0,0000	0,0067	0,0234
RMDv62	NaN	0,0260	0,9993	0,0048	0,7768	0,0004	0,1538	0,0000	0,0195	0,0475

Table 9: Normal distribution $p = 10$, $n = 100$, $\delta = 10$, $\lambda = 0.1$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,0149	0,0200	0,0202	0,0000	0,0445	0,0000	0,0896	0,0000	0,1500
MCD	NaN	0,1167	1,0000	0,0623	0,7000	0,0826	0,0500	0,2368	0,0000	0,3084
Adj.MCD	NaN	0,0970	1,0000	0,0431	0,7000	0,0666	0,0500	0,2096	0,0000	0,2767
Kurtosis	NaN	0,0609	0,9300	0,1091	0,4900	0,2686	0,9500	0,1012	0,9285	0,0806
RMDv1	NaN	0,0357	1,0000	0,0071	1,0000	0,0008	0,9914	0,0000	0,8007	0,0300
RMDv2	NaN	0,0357	1,0000	0,0071	1,0000	0,0010	0,9905	0,0000	0,8002	0,0300
RMDv3	NaN	0,0322	1,0000	0,0069	1,0000	0,0012	0,9900	0,0001	0,7908	0,0300
RMDv4	NaN	0,0239	1,0000	0,0049	1,0000	0,0007	1,0000	0,0000	0,9500	0,0300
RMDv5	NaN	0,0167	1,0000	0,0033	1,0000	0,0004	1,0000	0,0000	0,9500	0,0300
RMDv52	NaN	0,0228	1,0000	0,0047	1,0000	0,0005	1,0000	0,0000	0,9500	0,0300
RMDv6	NaN	0,0330	1,0000	0,0069	1,0000	0,0009	1,0000	0,0003	0,9196	0,0300
RMDv62	NaN	0,0248	1,0000	0,0054	1,0000	0,0007	1,0000	0,0000	0,9500	0,0300

Table 10: Normal distribution $p = 10$, $n = 100$, $\delta = 3$, $\lambda = 1$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,0171	0,2256	0,0104	0,0499	0,0109	0,0195	0,0105	0,0152	0,0122
MCD	NaN	0,1217	1,0000	0,0652	0,9015	0,0301	0,2128	0,1008	0,1222	0,1128
Adj.MCD	NaN	0,1027	1,0000	0,0461	0,8939	0,0180	0,1887	0,0837	0,1031	0,0924
Kurtosis	NaN	0,0690	1,0000	0,0874	0,6652	0,0947	0,3878	0,0625	0,4292	0,0451
RMDv1	NaN	0,0381	0,9987	0,0086	0,8175	0,0014	0,1391	0,0000	0,0013	0,0000
RMDv2	NaN	0,0378	0,9982	0,0081	0,8153	0,0012	0,1309	0,0000	0,0006	0,0000
RMDv3	NaN	0,0340	0,9976	0,0071	0,7983	0,0013	0,1249	0,0001	0,0016	0,0000
RMDv4	NaN	0,0251	0,9938	0,0049	0,8822	0,0012	0,3424	0,0000	0,0162	0,0000
RMDv5	NaN	0,0169	0,9909	0,0024	0,8198	0,0005	0,2098	0,0000	0,0040	0,0000
RMDv52	NaN	0,0234	0,9932	0,0044	0,8753	0,0011	0,3290	0,0000	0,0152	0,0000
RMDv6	NaN	0,0333	0,9895	0,0077	0,8232	0,0013	0,2042	0,0000	0,0054	0,0000
RMDv62	NaN	0,0261	0,9950	0,0047	0,8794	0,0013	0,3367	0,0000	0,0159	0,0000

Table 11: Normal distribution $p = 10$, $n = 100$, $\delta = 10$, $\lambda = 1$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,0171	0,2195	0,0095	0,0443	0,0123	0,0235	0,0113	0,0130	0,0123
MCD	NaN	0,1187	1,0000	0,0604	0,8686	0,0219	0,2638	0,0842	0,1320	0,1113
Adj.MCD	NaN	0,0995	1,0000	0,0416	0,8590	0,0116	0,2347	0,0706	0,1127	0,0918
Kurtosis	NaN	0,0784	0,9823	0,0993	0,7369	0,0958	0,5068	0,0638	0,9768	0,0465
RMDv1	NaN	0,0342	1,0000	0,0071	1,0000	0,0010	0,9964	0,0000	0,7706	0,0000
RMDv2	NaN	0,0341	1,0000	0,0071	1,0000	0,0012	0,9948	0,0000	0,7644	0,0000
RMDv3	NaN	0,0303	1,0000	0,0070	1,0000	0,0012	0,9914	0,0000	0,7490	0,0000
RMDv4	NaN	0,0229	1,0000	0,0036	1,0000	0,0003	1,0000	0,0000	0,9900	0,0000
RMDv5	NaN	0,0147	1,0000	0,0023	1,0000	0,0002	1,0000	0,0000	0,9889	0,0000
RMDv52	NaN	0,0220	1,0000	0,0033	1,0000	0,0002	1,0000	0,0000	0,9900	0,0000
RMDv6	NaN	0,0309	1,0000	0,0063	1,0000	0,0006	1,0000	0,0000	0,9695	0,0000
RMDv62	NaN	0,0241	1,0000	0,0034	1,0000	0,0005	1,0000	0,0000	0,9898	0,0000

Table 12: Normal distribution $p = 30$, $n = 500$, $\delta = 3$, $\lambda = 0.1$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,0209	0,0000	0,0444	0,0000	0,1067	0,0000	0,2335	0,0000	0,4251
MCD	NaN	0,0641	0,0000	0,0821	0,0000	0,1104	0,0000	0,1503	0,0000	0,2167
Adj.MCD	NaN	0,0408	0,0000	0,0556	0,0000	0,0809	0,0000	0,1164	0,0000	0,1772
Kurtosis	NaN	0,0079	0,9900	0,0119	0,1100	0,0591	0,6000	0,0566	1,0000	0,0221
RMDv1	NaN	0,0264	1,0000	0,0012	0,0322	0,0000	0,0000	0,0000	0,0000	0,0000
RMDv2	NaN	0,0261	1,0000	0,0012	0,0315	0,0000	0,0000	0,0000	0,0000	0,0000
RMDv3	NaN	0,0253	1,0000	0,0011	0,0285	0,0000	0,0000	0,0000	0,0000	0,0000
RMDv4	NaN	0,0041	1,0000	0,0001	0,5410	0,0000	0,2238	0,0131	0,0709	0,0147
RMDv5	NaN	0,0027	1,0000	0,0001	0,3144	0,0000	0,1283	0,0077	0,0323	0,0072
RMDv52	NaN	0,0041	1,0000	0,0001	0,5380	0,0000	0,2222	0,0129	0,0704	0,0144
RMDv6	NaN	0,0211	0,9900	0,0009	0,3656	0,0000	0,0000	0,0000	0,0100	0,0062
RMDv62	NaN	0,0042	1,0000	0,0001	0,5343	0,0000	0,2085	0,0087	0,0667	0,0090

Table 13: Normal distribution $p = 30$, $n = 500$, $\delta = 10$, $\lambda = 0.1$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,0211	0,0000	0,0424	0,0000	0,1090	0,0000	0,2313	0,0000	0,4240
MCD	NaN	0,0654	0,8900	0,0409	0,0000	0,1113	0,0000	0,1500	0,0000	0,2117
Adj.MCD	NaN	0,0418	0,8900	0,0168	0,0000	0,0815	0,0000	0,1161	0,0000	0,1720
Kurtosis	NaN	0,0084	1,0000	0,0085	0,0800	0,0586	0,6200	0,0544	1,0000	0,0221
RMDv1	NaN	0,0281	1,0000	0,0011	1,0000	0,0000	1,0000	0,0000	0,1432	0,0000
RMDv2	NaN	0,0279	1,0000	0,0011	1,0000	0,0000	1,0000	0,0000	0,1382	0,0000
RMDv3	NaN	0,0269	1,0000	0,0011	1,0000	0,0000	1,0000	0,0000	0,1299	0,0000
RMDv4	NaN	0,0050	1,0000	0,0001	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000
RMDv5	NaN	0,0029	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000
RMDv52	NaN	0,0050	1,0000	0,0001	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000
RMDv6	NaN	0,0225	1,0000	0,0008	1,0000	0,0000	1,0000	0,0000	0,9900	0,0000
RMDv62	NaN	0,0050	1,0000	0,0001	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000

Table 14: Normal distribution $p = 30$, $n = 500$, $\delta = 3$, $\lambda = 1$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,0215	0,1088	0,0161	0,0381	0,0162	0,0269	0,0163	0,0232	0,0167
MCD	NaN	0,0640	1,0000	0,0347	0,4080	0,0374	0,0833	0,0536	0,0719	0,0581
Adj.MCD	NaN	0,0405	1,0000	0,0111	0,3799	0,0223	0,0537	0,0329	0,0448	0,0370
Kurtosis	NaN	0,0077	0,8139	0,0076	0,0157	0,0063	0,0789	0,0079	0,6918	0,0162
RMDv1	NaN	0,0261	1,0000	0,0013	0,5739	0,0000	0,0003	0,0000	0,0000	0,0000
RMDv2	NaN	0,0258	1,0000	0,0013	0,5660	0,0000	0,0003	0,0000	0,0000	0,0000
RMDv3	NaN	0,0254	1,0000	0,0013	0,5566	0,0000	0,0003	0,0000	0,0000	0,0000
RMDv4	NaN	0,0049	1,0000	0,0001	0,8816	0,0000	0,0574	0,0000	0,0000	0,0000
RMDv5	NaN	0,0032	1,0000	0,0000	0,7537	0,0000	0,0173	0,0000	0,0000	0,0000
RMDv52	NaN	0,0049	1,0000	0,0001	0,8804	0,0000	0,0565	0,0000	0,0000	0,0000
RMDv6	NaN	0,0216	1,0000	0,0011	0,8824	0,0000	0,0528	0,0000	0,0000	0,0000
RMDv62	NaN	0,0050	1,0000	0,0001	0,8830	0,0000	0,0577	0,0000	0,0000	0,0000

Table 15: Normal distribution $p = 30$, $n = 500$, $\delta = 10$, $\lambda = 1$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,0210	0,1024	0,0169	0,0383	0,0158	0,0277	0,0161	0,0205	0,0192
MCD	NaN	0,0638	1,0000	0,0344	0,4229	0,0375	0,0823	0,0532	0,0675	0,0600
Adj.MCD	NaN	0,0403	1,0000	0,0111	0,3971	0,0213	0,0531	0,0321	0,0418	0,0383
Kurtosis	NaN	0,0082	0,8906	0,0089	0,0142	0,0061	0,0889	0,0079	0,6325	0,0169
RMDv1	NaN	0,0276	1,0000	0,0014	1,0000	0,0000	1,0000	0,0000	0,2447	0,0000
RMDv2	NaN	0,0277	1,0000	0,0013	1,0000	0,0000	1,0000	0,0000	0,2342	0,0000
RMDv3	NaN	0,0263	1,0000	0,0013	1,0000	0,0000	1,0000	0,0000	0,2189	0,0000
RMDv4	NaN	0,0051	1,0000	0,0001	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000
RMDv5	NaN	0,0034	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000
RMDv52	NaN	0,0050	1,0000	0,0001	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000
RMDv6	NaN	0,0230	1,0000	0,0009	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000
RMDv62	NaN	0,0052	1,0000	0,0001	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000

Table 16: Computational times in simulations with a Normal distribution with $p = 2$ and $n = 100$.

p=2	$\delta = 3$	$\lambda = 0.1$					$\delta = 3$	$\lambda = 1$				
Methods	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$	**Mean**	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$	**Mean**
Classic	0,01	0,00	0,00	0,00	0,00	**0,00**	0,01	0,00	0,00	0,00	0,00	**0,00**
MCD	1,36	1,33	1,28	1,24	1,26	**1,29**	1,48	1,64	1,31	1,37	1,30	**1,42**
Adj.MCD	1,25	1,32	1,27	1,27	1,25	**1,27**	1,26	1,29	1,25	1,25	1,40	**1,29**
Kurtosis	0,13	0,05	0,06	0,06	0,05	**0,07**	0,12	0,07	0,07	0,07	0,06	**0,08**
RMDv4	0,07	0,07	0,07	0,07	0,07	**0,07**	0,08	0,07	0,07	0,07	0,07	**0,07**
RMDv5	0,07	0,07	0,07	0,07	0,07	**0,07**	0,08	0,07	0,07	0,07	0,07	**0,07**
RMDv52	0,08	0,07	0,09	0,08	0,07	**0,08**	0,08	0,07	0,08	0,07	0,07	**0,08**
RMDv6	0,07	0,07	0,07	0,07	0,07	**0,07**	0,07	0,08	0,07	0,07	0,07	**0,07**
RMDv62	0,08	0,08	0,08	0,07	0,07	**0,08**	0,10	0,09	0,09	0,07	0,07	**0,09**
p=2	$\delta = 10$	$\lambda = 0.1$					$\delta = 10$	$\lambda = 1$				
Methods	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$	**Mean**	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$	**Mean**
Classic	0,01	0,00	0,00	0,00	0,00	**0,00**	0,01	0,00	0,00	0,00	0,00	**0,00**
MCD	1,57	1,38	1,40	1,50	1,35	**1,44**	1,57	1,47	1,27	1,22	1,30	**1,37**
Adj.MCD	1,26	1,31	1,34	1,59	1,29	**1,36**	1,34	1,24	1,42	1,35	1,41	**1,35**
Kurtosis	0,11	0,08	0,06	0,04	0,01	**0,06**	0,12	0,09	0,06	0,05	0,03	**0,07**
RMDv4	0,09	0,07	0,08	0,07	0,07	**0,08**	0,07	0,07	0,07	0,07	0,07	**0,07**
RMDv5	6,82	6,48	6,58	6,89	6,80	**6,71**	6,38	6,30	6,32	6,27	6,28	**6,31**
RMDv52	0,08	0,07	0,07	0,08	0,07	**0,08**	0,07	0,07	0,07	0,07	0,07	**0,07**
RMDv6	6,85	6,73	6,94	6,64	6,72	**6,78**	6,33	6,35	6,45	131,63	6,31	**31,42**
RMDv62	0,08	0,07	0,08	0,07	0,07	**0,07**	0,07	0,07	0,07	0,07	0,07	**0,07**

Table 17: Computational times in simulations with a Normal distribution with $p = 10$ and $n = 100$.

p=10	$\delta = 3$	$\lambda = 0.1$					$\delta = 3$	$\lambda = 1$				
Methods	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$	**Mean**	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$	**Mean**
Classic	0,01	0,00	0,00	0,00	0,00	**0,00**	0,01	0,00	0,00	0,00	0,00	**0,00**
MCD	2,03	1,89	2,08	1,89	1,86	**1,95**	1,97	1,86	1,84	1,93	1,93	**1,90**
Adj.MCD	1,89	1,87	1,89	1,94	1,96	**1,91**	1,84	1,89	1,89	1,92	1,91	**1,89**
Kurtosis	0,40	0,27	0,26	0,12	0,06	**0,22**	0,39	0,31	0,33	0,28	0,21	**0,30**
RMDv4	0,08	0,07	0,07	0,07	0,07	**0,07**	0,08	0,07	0,08	0,08	0,07	**0,08**
RMDv5	6,62	6,63	6,69	6,72	260,76	**57,48**	6,88	6,99	7,60	7,62	7,77	**7,37**
RMDv52	0,08	0,08	0,08	0,08	0,08	**0,08**	0,09	0,09	0,11	0,11	0,09	**0,10**
RMDv6	6,71	6,79	6,75	6,77	6,70	**6,75**	7,85	7,73	7,63	7,61	7,59	**7,68**
RMDv62	0,08	0,08	0,08	0,08	0,08	**0,08**	0,10	0,08	0,08	0,08	0,08	**0,09**
p=10	$\delta = 10$	$\lambda = 0.1$					$\delta = 10$	$\lambda = 1$				
Methods	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$	**Mean**	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$	**Mean**
Classic	0,01	0,00	0,00	0,00	0,00	**0,00**	0,01	0,00	0,00	0,00	0,00	**0,00**
MCD	1,97	1,92	1,89	1,90	1,92	**1,92**	3,14	2,63	2,49	2,62	2,72	**2,72**
Adj.MCD	1,92	1,89	1,83	1,89	1,89	**1,88**	2,95	2,57	2,66	3,41	3,03	**2,93**
Kurtosis	0,43	0,28	0,22	0,11	0,04	**0,22**	0,48	0,38	0,39	0,29	0,12	**0,33**
RMDv4	0,08	0,07	0,07	0,07	0,07	**0,07**	0,10	0,09	0,08	0,07	0,08	**0,08**
RMDv5	7,19	7,23	7,90	7,63	7,29	**7,45**	7,73	7,45	7,16	7,89	8,06	**7,66**
RMDv52	0,10	0,09	0,10	0,09	0,10	**0,09**	0,11	0,09	0,11	0,10	0,08	**0,10**
RMDv6	7,96	8,11	7,97	7,37	7,54	**7,79**	8,05	7,14	7,11	7,05	7,16	**7,30**
RMDv62	0,08	0,10	0,10	0,08	0,09	**0,09**	0,08	0,08	0,08	0,08	0,08	**0,08**

Table 18: Computational times in simulations with a Normal distribution with $p = 30$ and $n = 500$.

p=30	$\delta = 3$	$\lambda = 0.1$					$\delta = 3$	$\lambda = 1$				
Methods	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$	**Mean**	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$	**Mean**
Classic	0,02	0,01	0,01	0,01	0,01	**0,01**	0,25	0,00	0,00	0,00	0,00	**0,05**
MCD	10,03	9,17	9,25	9,19	9,16	**9,36**	2,77	2,11	2,14	2,10	2,09	**2,24**
Adj.MCD	136,56	10,11	9,05	9,07	8,93	**34,74**	2,18	2,08	2,11	2,10	2,09	**2,11**
Kurtosis	0,84	0,71	2,10	2,47	1,10	**1,44**	0,56	0,23	0,36	0,21	0,34	**0,34**
RMDv4	1,81	1,73	1,71	1,72	1,70	**1,73**	0,35	0,34	0,33	0,34	0,34	**0,34**
RMDv5	762,96	772,38	182,33	187,16	171,92	**415,35**	0,34	0,34	0,33	0,32	0,32	**0,33**
RMDv52	1,98	1,90	1,83	1,85	1,83	**1,88**	0,34	0,33	0,33	0,34	0,33	**0,34**
RMDv6	231,50	1604,35	377,23	170,20	176,32	**511,92**	0,32	0,32	0,32	0,32	0,32	**0,32**
RMDv62	1,74	1,89	1,92	1,84	1,88	**1,85**	0,33	0,33	0,33	0,33	0,34	**0,33**
p=30	$\delta = 10$	$\lambda = 0.1$					$\delta = 10$	$\lambda = 1$				
Methods	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$	**Mean**	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$	**Mean**
Classic	0,03	0,00	0,00	0,00	0,00	**0,01**	0,01	0,00	0,00	0,00	0,00	**0,00**
MCD	2,13	1,95	1,95	1,95	1,94	**1,98**	2,06	1,95	1,99	2,05	1,96	**2,00**
Adj.MCD	1,97	1,95	1,95	1,95	1,94	**1,95**	1,97	1,96	2,15	1,96	1,92	**1,99**
Kurtosis	0,28	0,29	0,54	0,67	0,18	**0,39**	0,50	0,27	0,25	0,18	0,15	**0,27**
RMDv4	0,33	0,31	0,31	0,32	0,31	**0,32**	0,32	0,31	0,31	0,31	0,33	**0,32**
RMDv5	0,32	0,32	0,31	0,31	0,31	**0,31**	0,31	0,31	0,31	0,31	0,32	**0,31**
RMDv52	0,33	0,32	0,32	0,32	0,32	**0,32**	0,33	0,33	0,32	0,32	0,32	**0,33**
RMDv6	0,31	0,31	0,30	0,31	0,31	**0,31**	0,31	0,31	0,31	0,31	0,31	**0,31**
RMDv62	0,32	0,32	0,32	0,32	0,32	**0,32**	0,32	0,32	0,33	0,32	0,32	**0,32**

Multivariate T distribution with 3 d.f.

The Tables 19-30 show the correct classification rates (c) and the false classification rates (f) for each method, corresponding to the simulations with multivariate T distribution with 3 degrees of freedom explained in Section 4.2.

Table 19: T (3 d.f.) distribution $p = 2$, $n = 100$, $\delta = 3$, $\lambda = 0.1$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,0500	0,0000	0,0421	0,0000	0,0471	0,0000	0,0511	0,0000	0,0624
MCD	NaN	0,0889	0,4908	0,0614	0,0417	0,1129	0,0003	0,1629	0,0005	0,2130
Adj.MCD	NaN	0,0724	0,4073	0,0517	0,0265	0,0944	0,0003	0,1381	0,0000	0,1836
Kurtosis	NaN	0,1074	0,9389	0,1117	0,4600	0,1480	0,0420	0,2101	0,0009	0,4588
RMDv1	NaN	0,1262	0,7858	0,0836	0,1520	0,0505	0,0015	0,0255	0,0000	0,0686
RMDv2	NaN	0,1264	0,7865	0,0844	0,1481	0,0512	0,0010	0,0263	0,0000	0,0668
RMDv3	NaN	0,1244	0,7879	0,0844	0,1563	0,0525	0,0010	0,0245	0,0000	0,0651
RMDv4	NaN	0,1235	0,7535	0,0828	0,1728	0,0520	0,0000	0,0263	0,0000	0,0767
RMDv5	NaN	0,1252	0,7555	0,0829	0,1713	0,0523	0,0000	0,0262	0,0000	0,0725
RMDv52	NaN	0,1253	0,7550	0,0822	0,1707	0,0524	0,0000	0,0265	0,0000	0,0743
RMDv6	NaN	0,1238	0,7680	0,0834	0,1671	0,0538	0,0000	0,0261	0,0000	0,0684
RMDv62	NaN	0,1245	0,7692	0,0830	0,1660	0,0534	0,0000	0,0260	0,0000	0,0720

Table 20: T (3 d.f.) distribution $p = 2$, $n = 100$, $\delta = 10$, $\lambda = 0.1$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,0497	0,6664	0,0264	0,0000	0,0292	0,0000	0,0383	0,0000	0,0555
MCD	NaN	0,0929	1,0000	0,0595	0,8700	0,0344	0,1896	0,1277	0,0002	0,2198
Adj.MCD	NaN	0,0762	1,0000	0,0437	0,8700	0,0253	0,1884	0,1097	0,0002	0,1892
Kurtosis	NaN	0,1052	1,0000	0,1108	0,9500	0,1061	0,9683	0,1023	0,7802	0,2283
RMDv1	NaN	0,1274	1,0000	0,0910	1,0000	0,0493	1,0000	0,0240	0,6167	0,0075
RMDv2	NaN	0,1278	1,0000	0,0911	1,0000	0,0501	1,0000	0,0243	0,6055	0,0079
RMDv3	NaN	0,1273	1,0000	0,0906	1,0000	0,0506	1,0000	0,0244	0,6029	0,0074
RMDv4	NaN	0,1263	1,0000	0,0899	1,0000	0,0497	1,0000	0,0243	0,6743	0,0086
RMDv5	NaN	0,1263	1,0000	0,0904	1,0000	0,0498	1,0000	0,0249	0,6612	0,0084
RMDv52	NaN	0,1260	1,0000	0,0899	1,0000	0,0490	1,0000	0,0247	0,6627	0,0084
RMDv6	NaN	0,1259	1,0000	0,0899	1,0000	0,0509	1,0000	0,0243	0,6423	0,0079
RMDv62	NaN	0,1260	1,0000	0,0911	1,0000	0,0508	1,0000	0,0249	0,6547	0,0081

Table 21: T (3 d.f.) distribution $p = 2$, $n = 100$, $\delta = 3$, $\lambda = 1$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,0486	0,1572	0,0379	0,0744	0,0349	0,0485	0,0337	0,0413	0,0323
MCD	NaN	0,1018	0,5953	0,0605	0,2735	0,0514	0,1216	0,0663	0,0897	0,0752
Adj.MCD	NaN	0,0847	0,5025	0,0509	0,2163	0,0417	0,0954	0,0531	0,0698	0,0594
Kurtosis	NaN	0,1144	0,7637	0,1046	0,4184	0,0802	0,2209	0,0733	0,0843	0,0580
RMDv1	NaN	0,1255	0,7627	0,0970	0,4428	0,0627	0,1599	0,0366	0,0534	0,0287
RMDv2	NaN	0,1242	0,7610	0,0969	0,4392	0,0629	0,1593	0,0358	0,0535	0,0272
RMDv3	NaN	0,1245	0,7563	0,0961	0,4284	0,0624	0,1589	0,0376	0,0541	0,0277
RMDv4	NaN	0,1235	0,7578	0,0966	0,4350	0,0614	0,1636	0,0379	0,0553	0,0275
RMDv5	NaN	0,1234	0,7558	0,0957	0,4368	0,0614	0,1576	0,0367	0,0557	0,0270
RMDv52	NaN	0,1239	0,7579	0,0960	0,4366	0,0615	0,1604	0,0367	0,0557	0,0272
RMDv6	NaN	0,1234	0,7610	0,0967	0,4324	0,0622	0,1593	0,0367	0,0540	0,0274
RMDv62	NaN	0,1228	0,7605	0,0964	0,4321	0,0621	0,1614	0,0371	0,0531	0,0277

Table 22: T (3 d.f.) distribution $p = 2$, $n = 100$, $\delta = 10$, $\lambda = 1$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,0502	0,6613	0,0231	0,0959	0,0235	0,0366	0,0233	0,0338	0,0240
MCD	NaN	0,0909	1,0000	0,0554	0,9285	0,0197	0,1807	0,0485	0,1006	0,0789
Adj.MCD	NaN	0,0739	1,0000	0,0414	0,9194	0,0136	0,1632	0,0393	0,0833	0,0627
Kurtosis	NaN	0,1074	1,0000	0,1125	1,0000	0,1275	0,9902	0,1174	0,9038	0,1036
RMDv1	NaN	0,1284	1,0000	0,0886	1,0000	0,0556	0,9807	0,0250	0,7214	0,0088
RMDv2	NaN	0,1283	1,0000	0,0888	1,0000	0,0559	0,9804	0,0247	0,7188	0,0091
RMDv3	NaN	0,1275	1,0000	0,0875	1,0000	0,0553	0,9806	0,0239	0,7131	0,0088
RMDv4	NaN	0,1262	1,0000	0,0859	1,0000	0,0560	0,9783	0,0255	0,7326	0,0095
RMDv5	NaN	0,1254	1,0000	0,0868	1,0000	0,0560	0,9780	0,0248	0,7283	0,0095
RMDv52	NaN	0,1253	1,0000	0,0863	1,0000	0,0558	0,9780	0,0249	0,7290	0,0095
RMDv6	NaN	0,1255	1,0000	0,0869	1,0000	0,0539	0,9806	0,0245	0,7268	0,0088
RMDv62	NaN	0,1251	1,0000	0,0865	1,0000	0,0551	0,9806	0,0247	0,7286	0,0095

Table 23: T (3 d.f.) distribution $p = 10$, $n = 100$, $\delta = 3$, $\lambda = 0.1$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,0892	0,0000	0,0885	0,0005	0,1182	0,0000	0,1670	0,0000	0,2372
MCD	NaN	0,1804	0,1900	0,1965	0,0005	0,2409	0,0000	0,2965	0,0003	0,3723
Adj.MCD	NaN	0,1619	0,1900	0,1759	0,0005	0,2188	0,0000	0,2712	0,0003	0,3497
Kurtosis	NaN	0,3237	0,3400	0,3704	0,0305	0,5124	0,0118	0,6084	0,0000	0,7485
RMDv1	NaN	0,3264	0,9875	0,2024	0,2948	0,1097	0,0000	0,0465	0,0000	0,1886
RMDv2	NaN	0,3272	0,9850	0,2053	0,2840	0,1104	0,0000	0,0469	0,0000	0,1815
RMDv3	NaN	0,3232	0,9800	0,2004	0,2554	0,1090	0,0000	0,0447	0,0000	0,1649
RMDv4	NaN	0,2602	0,9736	0,1591	0,3772	0,0955	0,0219	0,0638	0,0497	0,2371
RMDv5	NaN	0,2447	0,9545	0,1508	0,2518	0,0905	0,0077	0,0529	0,0022	0,2066
RMDv52	NaN	0,2583	0,9707	0,1568	0,3625	0,0938	0,0208	0,0619	0,0381	0,2326
RMDv6	NaN	0,3153	0,9500	0,1990	0,4235	0,1131	0,0100	0,0475	0,0045	0,2176
RMDv62	NaN	0,2674	0,9700	0,1636	0,3875	0,0979	0,0120	0,0595	0,0262	0,2276

Table 24: T (3 d.f.) distribution $p = 10$, $n = 100$, $\delta = 10$, $\lambda = 0.1$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,0840	0,0100	0,0853	0,0000	0,1180	0,0000	0,1604	0,0002	0,2332
MCD	NaN	0,1718	0,9600	0,1250	0,1200	0,2301	0,0003	0,3018	0,0002	0,3714
Adj.MCD	NaN	0,1533	0,9600	0,1062	0,1200	0,2072	0,0003	0,2761	0,0002	0,3471
Kurtosis	NaN	0,3226	0,5100	0,3574	0,2406	0,4420	0,7600	0,2460	0,8572	0,1284
RMDv1	NaN	0,3143	1,0000	0,2033	1,0000	0,1045	0,9800	0,0432	0,4732	0,0392
RMDv2	NaN	0,3139	1,0000	0,2050	1,0000	0,1061	0,9800	0,0434	0,4700	0,0324
RMDv3	NaN	0,3085	1,0000	0,2003	1,0000	0,1033	0,9708	0,0440	0,4351	0,0212
RMDv4	NaN	0,2439	1,0000	0,1602	1,0000	0,0898	1,0000	0,0428	0,8442	0,0340
RMDv5	NaN	0,2275	1,0000	0,1481	1,0000	0,0860	1,0000	0,0422	0,8132	0,0340
RMDv52	NaN	0,2419	1,0000	0,1576	1,0000	0,0886	1,0000	0,0426	0,8396	0,0338
RMDv6	NaN	0,3001	1,0000	0,1967	1,0000	0,1079	1,0000	0,0482	0,7107	0,0463
RMDv62	NaN	0,2517	1,0000	0,1630	1,0000	0,0932	1,0000	0,0447	0,8338	0,0346

Table 25: T (3 d.f.) distribution $p = 10$, $n = 100$, $\delta = 3$, $\lambda = 1$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,0874	0,1994	0,0725	0,1081	0,0751	0,0855	0,0770	0,0889	0,0771
MCD	NaN	0,1761	0,9360	0,1207	0,3079	0,1514	0,1877	0,1650	0,1809	0,1718
Adj.MCD	NaN	0,1574	0,9235	0,1034	0,2797	0,1357	0,1674	0,1472	0,1647	0,1522
Kurtosis	NaN	0,3036	0,9386	0,2842	0,4801	0,2845	0,3355	0,2784	0,3015	0,2631
RMDv1	NaN	0,3138	0,9977	0,2094	0,7371	0,1289	0,2555	0,0761	0,0795	0,0569
RMDv2	NaN	0,3146	0,9968	0,2100	0,7312	0,1302	0,2440	0,0757	0,0779	0,0566
RMDv3	NaN	0,3087	0,9948	0,2037	0,7093	0,1282	0,2334	0,0749	0,0783	0,0553
RMDv4	NaN	0,2470	0,9876	0,1601	0,7808	0,1097	0,4010	0,0688	0,1060	0,0460
RMDv5	NaN	0,2351	0,9740	0,1527	0,6964	0,1038	0,3246	0,0651	0,0842	0,0428
RMDv52	NaN	0,2439	0,9848	0,1585	0,7717	0,1085	0,3886	0,0684	0,1025	0,0446
RMDv6	NaN	0,2967	0,9765	0,1951	0,7592	0,1303	0,3823	0,0786	0,1029	0,0511
RMDv62	NaN	0,2542	0,9876	0,1656	0,7832	0,1127	0,4069	0,0705	0,1060	0,0457

Table 26: T (3 d.f.) distribution $p = 10$, $n = 100$, $\delta = 10$, $\lambda = 1$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,0841	0,2488	0,0707	0,1100	0,0718	0,0934	0,0778	0,0806	0,0739
MCD	NaN	0,1750	1,0000	0,1116	0,7132	0,0931	0,2752	0,1539	0,1862	0,1720
Adj.MCD	NaN	0,1563	1,0000	0,0936	0,7044	0,0806	0,2561	0,1366	0,1650	0,1544
Kurtosis	NaN	0,3075	0,9371	0,2886	0,5491	0,2696	0,6589	0,2114	0,8843	0,0963
RMDv1	NaN	0,3190	1,0000	0,1955	1,0000	0,1006	0,9931	0,0457	0,4997	0,0117
RMDv2	NaN	0,3208	1,0000	0,1959	1,0000	0,1003	0,9914	0,0462	0,4865	0,0124
RMDv3	NaN	0,3152	1,0000	0,1935	1,0000	0,1007	0,9874	0,0470	0,4491	0,0111
RMDv4	NaN	0,2569	1,0000	0,1550	1,0000	0,0818	1,0000	0,0425	0,8809	0,0174
RMDv5	NaN	0,2389	1,0000	0,1454	1,0000	0,0782	1,0000	0,0413	0,8435	0,0162
RMDv52	NaN	0,2539	1,0000	0,1522	1,0000	0,0806	1,0000	0,0423	0,8790	0,0174
RMDv6	NaN	0,3062	1,0000	0,1908	1,0000	0,1016	1,0000	0,0524	0,6903	0,0198
RMDv62	NaN	0,2617	1,0000	0,1610	1,0000	0,0854	1,0000	0,0443	0,8757	0,0177

Table 27: T (3 d.f.) distribution $p = 30$, $n = 500$, $\delta = 3$, $\lambda = 0.1$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,1134	0,0002	0,1401	0,0000	0,1895	0,0001	0,2665	0,0002	0,3707
MCD	NaN	0,1569	0,0002	0,1808	0,0000	0,2097	0,0001	0,2577	0,0001	0,3249
Adj.MCD	NaN	0,1340	0,0002	0,1553	0,0000	0,1809	0,0001	0,2244	0,0001	0,2870
Kurtosis	NaN	0,2088	0,1702	0,3218	0,1501	0,4977	0,1397	0,5821	0,0000	0,7826
RMDv1	NaN	0,5664	0,9903	0,3538	0,0000	0,1595	0,0000	0,0497	0,0000	0,0610
RMDv2	NaN	0,5674	0,9900	0,3553	0,0000	0,1599	0,0000	0,0496	0,0000	0,0602
RMDv3	NaN	0,5653	0,9900	0,3531	0,0000	0,1600	0,0000	0,0497	0,0000	0,0567
RMDv4	NaN	0,3699	0,9749	0,2274	0,2029	0,1376	0,1246	0,0937	0,0200	0,0915
RMDv5	NaN	0,3535	0,9578	0,2151	0,1031	0,1303	0,0920	0,0821	0,0066	0,0918
RMDv52	NaN	0,3692	0,9746	0,2272	0,1996	0,1374	0,1239	0,0933	0,0198	0,0913
RMDv6	NaN	0,5170	0,9990	0,3260	0,0580	0,1624	0,0101	0,0664	0,0000	0,0794
RMDv62	NaN	0,3721	0,9757	0,2289	0,1840	0,1357	0,1240	0,0922	0,0272	0,0885

Table 28: T (3 d.f.) distribution $p = 30$, $n = 500$, $\delta = 10$, $\lambda = 0.1$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,1165	0,0000	0,1375	0,0001	0,1898	0,0000	0,2637	0,0001	0,3760
MCD	NaN	0,1565	0,0100	0,1799	0,0001	0,2147	0,0000	0,2586	0,0001	0,3215
Adj.MCD	NaN	0,1337	0,0100	0,1546	0,0001	0,1859	0,0000	0,2255	0,0001	0,2831
Kurtosis	NaN	0,2088	0,3300	0,2991	0,2402	0,4917	0,9900	0,1662	0,9900	0,0603
RMDv1	NaN	0,5637	1,0000	0,3548	1,0000	0,1617	0,9860	0,0515	0,0186	0,0100
RMDv2	NaN	0,5638	1,0000	0,3559	1,0000	0,1627	0,9785	0,0515	0,0155	0,0101
RMDv3	NaN	0,5625	1,0000	0,3533	1,0000	0,1625	0,9713	0,0520	0,0114	0,0099
RMDv4	NaN	0,3692	1,0000	0,2230	1,0000	0,1117	1,0000	0,0492	0,9100	0,0130
RMDv5	NaN	0,3523	1,0000	0,2112	1,0000	0,1066	1,0000	0,0476	0,9010	0,0128
RMDv52	NaN	0,3688	1,0000	0,2227	1,0000	0,1115	1,0000	0,0491	0,9100	0,0130
RMDv6	NaN	0,5228	1,0000	0,3271	1,0000	0,1626	1,0000	0,0651	0,8683	0,0160
RMDv62	NaN	0,3714	1,0000	0,2247	1,0000	0,1123	1,0000	0,0494	0,9100	0,0129

Table 29: T (3 d.f.) distribution $p = 30$, $n = 500$, $\delta = 3$, $\lambda = 1$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,1144	0,1541	0,1083	0,1260	0,1076	0,1196	0,1074	0,1180	0,1091
MCD	NaN	0,1566	0,5640	0,1293	0,1703	0,1506	0,1644	0,1513	0,1619	0,1529
Adj.MCD	NaN	0,1335	0,5462	0,1072	0,1422	0,1287	0,1385	0,1295	0,1383	0,1303
Kurtosis	NaN	0,2137	0,4061	0,1875	0,2242	0,2013	0,2774	0,1943	0,2960	0,1767
RMDv1	NaN	0,5641	1,0000	0,3642	0,8149	0,2058	0,1760	0,0995	0,0819	0,0676
RMDv2	NaN	0,5640	1,0000	0,3649	0,8129	0,2068	0,1741	0,0999	0,0812	0,0678
RMDv3	NaN	0,5623	1,0000	0,3635	0,8058	0,2048	0,1743	0,0996	0,0810	0,0672
RMDv4	NaN	0,3718	1,0000	0,2339	0,8669	0,1455	0,2708	0,0780	0,0838	0,0477
RMDv5	NaN	0,3537	0,9994	0,2208	0,7997	0,1381	0,2231	0,0748	0,0746	0,0450
RMDv52	NaN	0,3713	1,0000	0,2334	0,8653	0,1453	0,2697	0,0779	0,0835	0,0476
RMDv6	NaN	0,5169	0,9984	0,3344	0,9340	0,2039	0,3434	0,0999	0,0998	0,0562
RMDv62	NaN	0,3742	1,0000	0,2355	0,8696	0,1463	0,2728	0,0785	0,0838	0,0479

Table 30: T (3 d.f.) distribution $p = 30$, $n = 500$, $\delta = 10$, $\lambda = 1$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,1150	0,1739	0,1078	0,1319	0,1083	0,1215	0,1102	0,1145	0,1078
MCD	NaN	0,1580	1,0000	0,1033	0,2667	0,1393	0,1646	0,1520	0,1594	0,1529
Adj.MCD	NaN	0,1351	1,0000	0,0808	0,2414	0,1182	0,1413	0,1292	0,1361	0,1299
Kurtosis	NaN	0,2121	0,5059	0,1933	0,2058	0,1834	0,6237	0,1732	0,9999	0,0854
RMDv1	NaN	0,5628	1,0000	0,3511	1,0000	0,1636	0,9999	0,0531	0,1451	0,0121
RMDv2	NaN	0,5642	1,0000	0,3524	1,0000	0,1648	0,9999	0,0534	0,1375	0,0122
RMDv3	NaN	0,5616	1,0000	0,3512	1,0000	0,1638	1,0000	0,0527	0,1215	0,0122
RMDv4	NaN	0,3723	1,0000	0,2219	1,0000	0,1145	1,0000	0,0480	0,9928	0,0159
RMDv5	NaN	0,3540	1,0000	0,2095	1,0000	0,1095	1,0000	0,0465	0,9877	0,0154
RMDv52	NaN	0,3717	1,0000	0,2217	1,0000	0,1142	1,0000	0,0480	0,9926	0,0159
RMDv6	NaN	0,5197	1,0000	0,3268	1,0000	0,1657	1,0000	0,0663	0,9563	0,0188
RMDv62	NaN	0,3747	1,0000	0,2242	1,0000	0,1154	1,0000	0,0485	0,9925	0,0159

Multivariate Exponential distribution

The following tables show the correct classification rates (c) and the false classification rates (f) for each method, corresponding to the simulations with multivariate Exponential distribution explained in Section 4.3.

Table 31: Exponential distribution $p = 2$, $n = 100$, $\delta = 3$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,0540	0,3848	0,0193	0,2667	0,0096	0,2073	0,0041	0,1687	0,0016
MCD	NaN	0,1264	0,5930	0,0897	0,5197	0,0592	0,4429	0,0363	0,3746	0,0159
Adj.MCD	NaN	0,1090	0,5623	0,0738	0,4819	0,0489	0,4111	0,0278	0,3381	0,0126
Kurtosis	NaN	0,2261	0,7338	0,2168	0,7065	0,1942	0,6678	0,1697	0,6633	0,1524
RMDv1	NaN	0,1510	0,6500	0,1243	0,5779	0,0931	0,5221	0,0654	0,4907	0,0389
RMDv2	NaN	0,1511	0,6553	0,1222	0,5791	0,0930	0,5224	0,0646	0,4892	0,0384
RMDv3	NaN	0,1508	0,6565	0,1219	0,5806	0,0893	0,5194	0,0620	0,4863	0,0379
RMDv4	NaN	0,1524	0,6571	0,1262	0,5815	0,0964	0,5236	0,0692	0,4910	0,0447
RMDv5	NaN	0,1516	0,6573	0,1257	0,5818	0,0953	0,5227	0,0679	0,4892	0,0442
RMDv52	NaN	0,1525	0,6560	0,1262	0,5823	0,0958	0,5229	0,0675	0,4905	0,0454
RMDv6	NaN	0,1484	0,6561	0,1247	0,5795	0,0915	0,5220	0,0641	0,4906	0,0414
RMDv62	NaN	0,1505	0,6555	0,1266	0,5835	0,0929	0,5191	0,0665	0,4907	0,0430

Table 32: Exponential distribution $p = 2$, $n = 100$, $\delta = 10$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,0563	0,5820	0,0018	0,3719	0,0000	0,2460	0,0000	0,1873	0,0000
MCD	NaN	0,1243	0,9231	0,0742	0,8609	0,0272	0,6995	0,0041	0,4954	0,0003
Adj.MCD	NaN	0,1070	0,9119	0,0581	0,8386	0,0189	0,6604	0,0021	0,4592	0,0002
Kurtosis	NaN	0,2194	0,9554	0,2092	0,9693	0,2185	0,9618	0,1881	0,9423	0,1353
RMDv1	NaN	0,1553	0,9344	0,1004	0,9168	0,0657	0,8798	0,0277	0,8123	0,0115
RMDv2	NaN	0,1531	0,9352	0,1007	0,9140	0,0649	0,8795	0,0275	0,8147	0,0112
RMDv3	NaN	0,1531	0,9357	0,0996	0,9139	0,0634	0,8785	0,0261	0,8148	0,0124
RMDv4	NaN	0,1613	0,9386	0,1039	0,9157	0,0668	0,8851	0,0322	0,8157	0,0139
RMDv5	NaN	0,1594	0,9383	0,1024	0,9144	0,0655	0,8847	0,0302	0,8155	0,0134
RMDv52	NaN	0,1597	0,9386	0,1028	0,9161	0,0664	0,8853	0,0312	0,8157	0,0133
RMDv6	NaN	0,1575	0,9379	0,1023	0,9166	0,0630	0,8841	0,0284	0,8146	0,0147
RMDv62	NaN	0,1590	0,9379	0,1035	0,9160	0,0641	0,8851	0,0296	0,8136	0,0141

Table 33: Exponential distribution $p = 10$, $n = 100$, $\delta = 3$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,0865	0,8234	0,0300	0,6800	0,0111	0,5218	0,0046	0,4113	0,0014
MCD	NaN	0,1885	0,9611	0,1222	0,9091	0,0526	0,7193	0,0157	0,5713	0,0043
Adj.MCD	NaN	0,1699	0,9495	0,1052	0,8890	0,0411	0,6911	0,0122	0,5423	0,0030
Kurtosis	NaN	0,3641	0,9844	0,3267	0,9784	0,2786	0,9537	0,1914	0,8930	0,1195
RMDv1	NaN	0,3855	0,9904	0,2918	0,9799	0,1976	0,9571	0,1153	0,9208	0,0763
RMDv2	NaN	0,3835	0,9890	0,2869	0,9802	0,1939	0,9560	0,1129	0,9213	0,0734
RMDv3	NaN	0,3711	0,9894	0,2721	0,9789	0,1856	0,9548	0,1064	0,9143	0,0639
RMDv4	NaN	0,4434	0,9904	0,3554	0,9801	0,2710	0,9680	0,1730	0,9376	0,1367
RMDv5	NaN	0,4100	0,9878	0,3231	0,9785	0,2451	0,9637	0,1525	0,9336	0,1153
RMDv52	NaN	0,4387	0,9904	0,3506	0,9805	0,2667	0,9671	0,1697	0,9371	0,1327
RMDv6	NaN	0,4410	0,9883	0,3263	0,9827	0,2525	0,9673	0,1714	0,9427	0,1219
RMDv62	NaN	0,4531	0,9904	0,3605	0,9816	0,2778	0,9699	0,1828	0,9398	0,1423

Table 34: Exponential distribution $p = 10$, $n = 100$, $\delta = 10$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,0867	0,9866	0,0080	0,8876	0,0000	0,6766	0,0000	0,5008	0,0000
MCD	NaN	0,1901	1,0000	0,1186	0,9837	0,0405	0,7814	0,0021	0,5783	0,0000
Adj.MCD	NaN	0,1714	1,0000	0,0997	0,9837	0,0308	0,7801	0,0015	0,5718	0,0000
Kurtosis	NaN	0,3720	1,0000	0,3290	1,0000	0,2663	0,9997	0,1835	0,9830	0,0740
RMDv1	NaN	0,3936	1,0000	0,2194	1,0000	0,1089	1,0000	0,0484	0,9996	0,0102
RMDv2	NaN	0,3914	1,0000	0,2198	1,0000	0,1061	1,0000	0,0463	0,9996	0,0090
RMDv3	NaN	0,3786	1,0000	0,2098	1,0000	0,0991	1,0000	0,0428	0,9993	0,0076
RMDv4	NaN	0,4445	1,0000	0,2997	1,0000	0,1739	1,0000	0,0956	0,9998	0,0338
RMDv5	NaN	0,4146	1,0000	0,2664	1,0000	0,1454	1,0000	0,0754	0,9998	0,0250
RMDv52	NaN	0,4398	1,0000	0,2953	1,0000	0,1683	1,0000	0,0912	0,9998	0,0325
RMDv6	NaN	0,4478	1,0000	0,2802	1,0000	0,1421	1,0000	0,0753	0,9998	0,0259
RMDv62	NaN	0,4572	1,0000	0,3074	1,0000	0,1752	1,0000	0,1011	0,9998	0,0363

Table 35: Exponential distribution $p = 30$, $n = 500$, $\delta = 3$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,1134	0,9866	0,0138	0,9327	0,0019	0,8150	0,0003	0,6697	0,0001
MCD	NaN	0,1410	0,9994	0,0850	0,9974	0,0262	0,7635	0,0002	0,5322	0,0000
Adj.MCD	NaN	0,1177	0,9987	0,0608	0,9939	0,0116	0,7086	0,0001	0,4797	0,0000
Kurtosis	NaN	0,1068	0,9977	0,1321	0,9945	0,1513	0,9918	0,1348	0,9732	0,0854
RMDv1	NaN	0,6393	1,0000	0,4476	1,0000	0,2830	0,9997	0,1485	0,9988	0,0649
RMDv2	NaN	0,6385	1,0000	0,4472	1,0000	0,2814	0,9997	0,1472	0,9989	0,0641
RMDv3	NaN	0,6339	1,0000	0,4411	1,0000	0,2777	0,9996	0,1442	0,9988	0,0618
RMDv4	NaN	0,6776	0,9998	0,5173	1,0000	0,3961	0,9999	0,2609	0,9992	0,1477
RMDv5	NaN	0,6476	0,9998	0,4829	1,0000	0,3615	0,9999	0,2361	0,9990	0,1282
RMDv52	NaN	0,6768	0,9998	0,5164	1,0000	0,3952	0,9999	0,2600	0,9992	0,1470
RMDv6	NaN	0,7193	0,9998	0,5736	1,0000	0,4427	0,9999	0,3247	0,9998	0,1895
RMDv62	NaN	0,6800	0,9998	0,5200	1,0000	0,3993	0,9999	0,2642	0,9992	0,1495

Table 36: Exponential distribution $p = 30$, $n = 500$, $\delta = 10$

	$\alpha = 0$		$\alpha = 0.1$		$\alpha = 0.2$		$\alpha = 0.3$		$\alpha = 0.4$	
Methods	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**	**c**	**f**
Classic	NaN	0,1130	1,0000	0,0000	0,9915	0,0000	0,9284	0,0000	0,7926	0,0000
MCD	NaN	0,1452	1,0000	0,0867	1,0000	0,0261	0,7860	0,0000	0,5630	0,0000
Adj.MCD	NaN	0,1219	1,0000	0,0627	1,0000	0,0113	0,7754	0,0000	0,5249	0,0000
Kurtosis	NaN	0,1374	1,0000	0,1297	1,0000	0,1298	1,0000	0,1214	1,0000	0,0875
RMDv1	NaN	0,6436	1,0000	0,3515	1,0000	0,1340	1,0000	0,0287	1,0000	0,0034
RMDv2	NaN	0,6426	1,0000	0,3497	1,0000	0,1331	1,0000	0,0281	1,0000	0,0033
RMDv3	NaN	0,6373	1,0000	0,3450	1,0000	0,1300	1,0000	0,0269	1,0000	0,0029
RMDv4	NaN	0,6842	1,0000	0,4455	1,0000	0,2460	1,0000	0,0900	1,0000	0,0223
RMDv5	NaN	0,6536	1,0000	0,4052	1,0000	0,2169	1,0000	0,0746	1,0000	0,0178
RMDv52	NaN	0,6834	1,0000	0,4445	1,0000	0,2450	1,0000	0,0896	1,0000	0,0221
RMDv6	NaN	0,7206	1,0000	0,4809	1,0000	0,2871	1,0000	0,1125	1,0000	0,0283
RMDv62	NaN	0,6871	1,0000	0,4490	1,0000	0,2481	1,0000	0,0912	1,0000	0,0230

Appendix C Figures

Simulation study

The following figures illustrates the results in the tables from Appendix B.

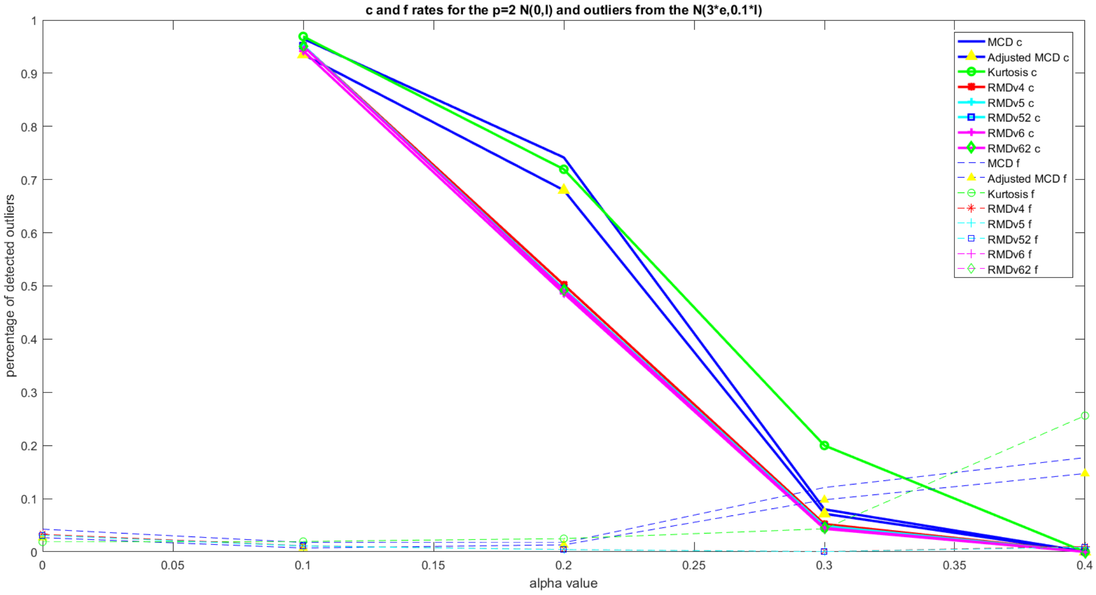

Figure 1: Percentages of detected outliers (c and f) for each method.

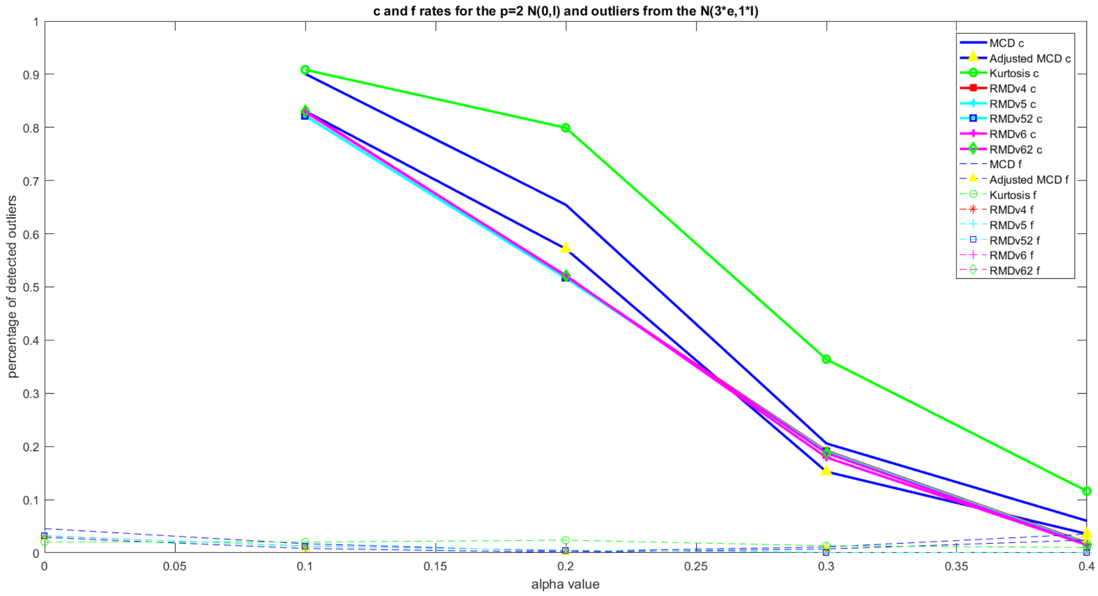

Figure 2: Percentages of detected outliers (c and f) for each method.

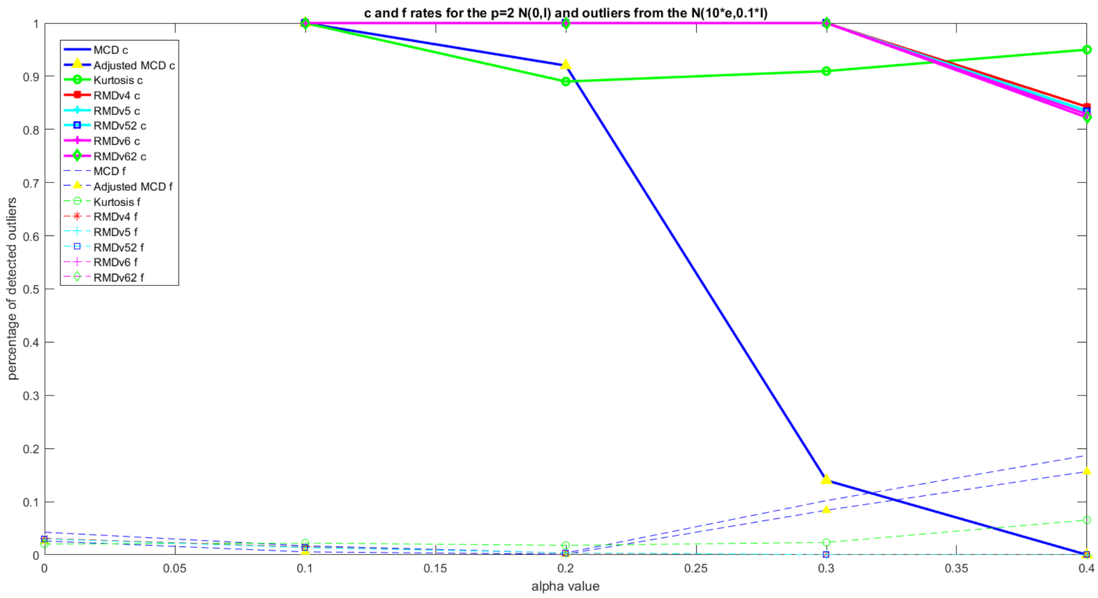

Figure 3: Percentages of detected outliers (c and f) for each method.

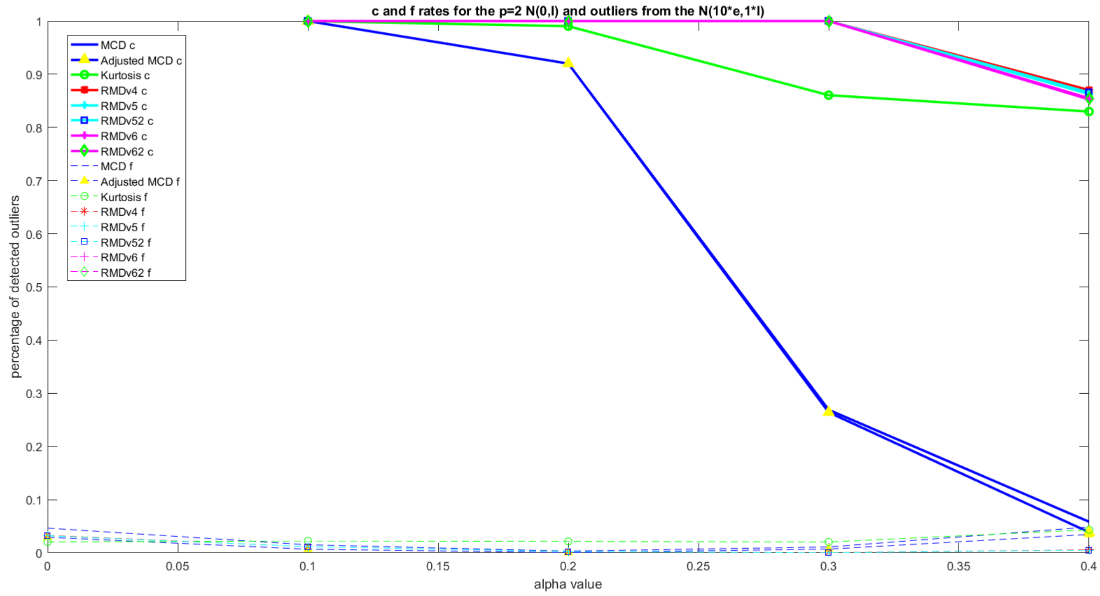

Figure 4: Percentages of detected outliers (c and f) for each method.

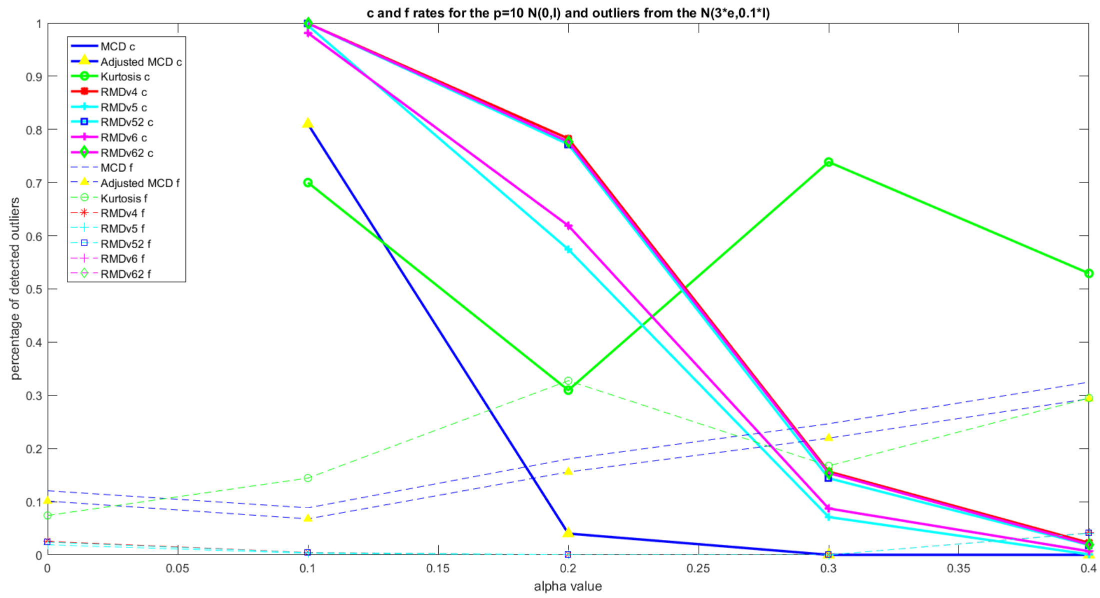

Figure 5: Percentages of detected outliers (c and f) for each method.

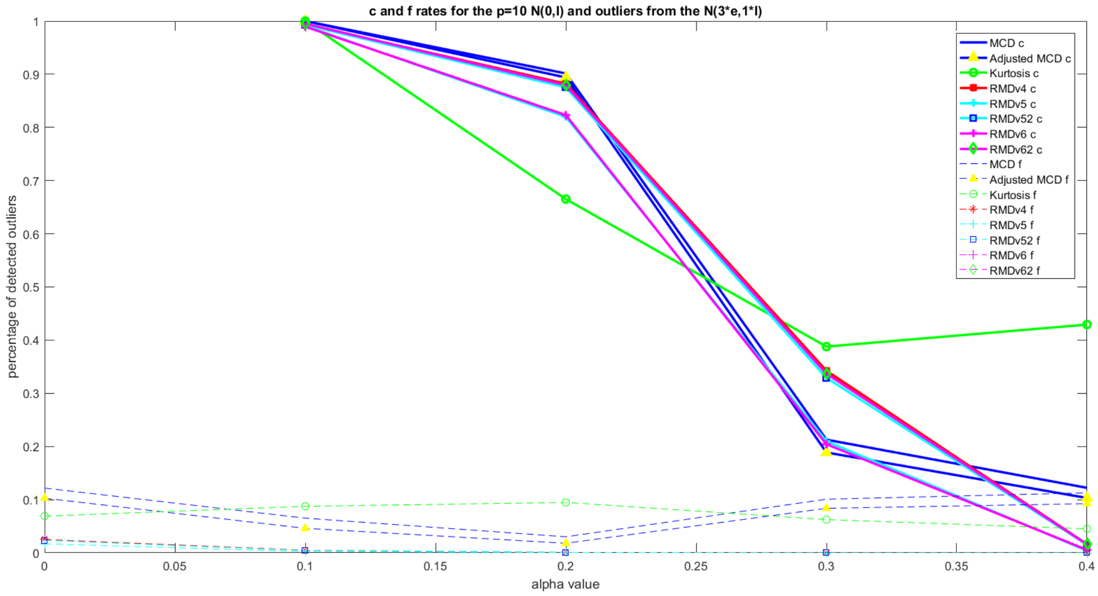

Figure 6: Percentages of detected outliers (c and f) for each method.

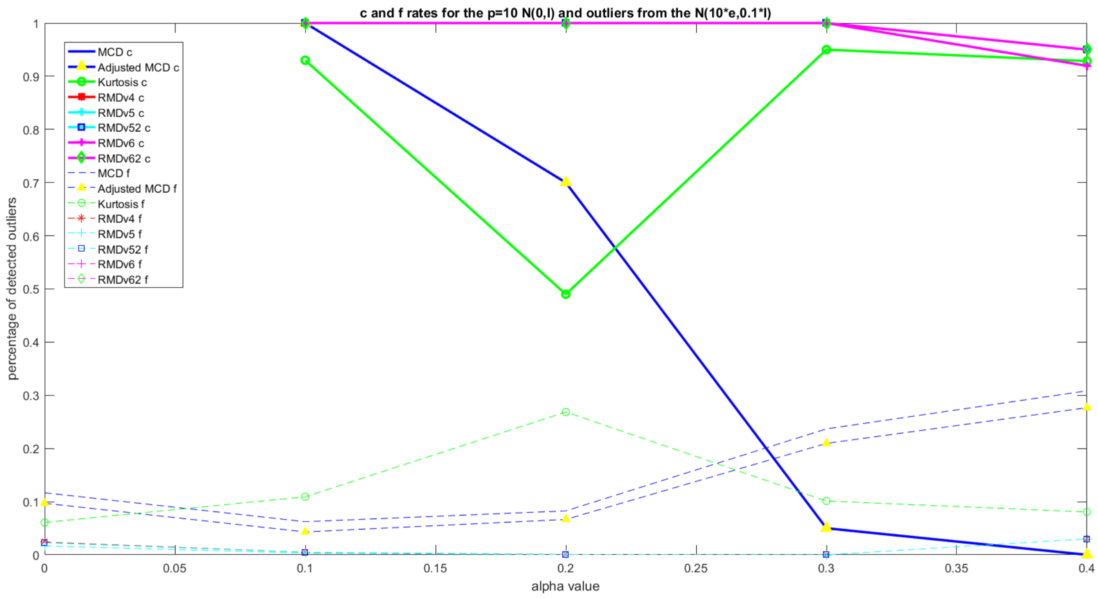

Figure 7: Percentages of detected outliers (c and f) for each method.

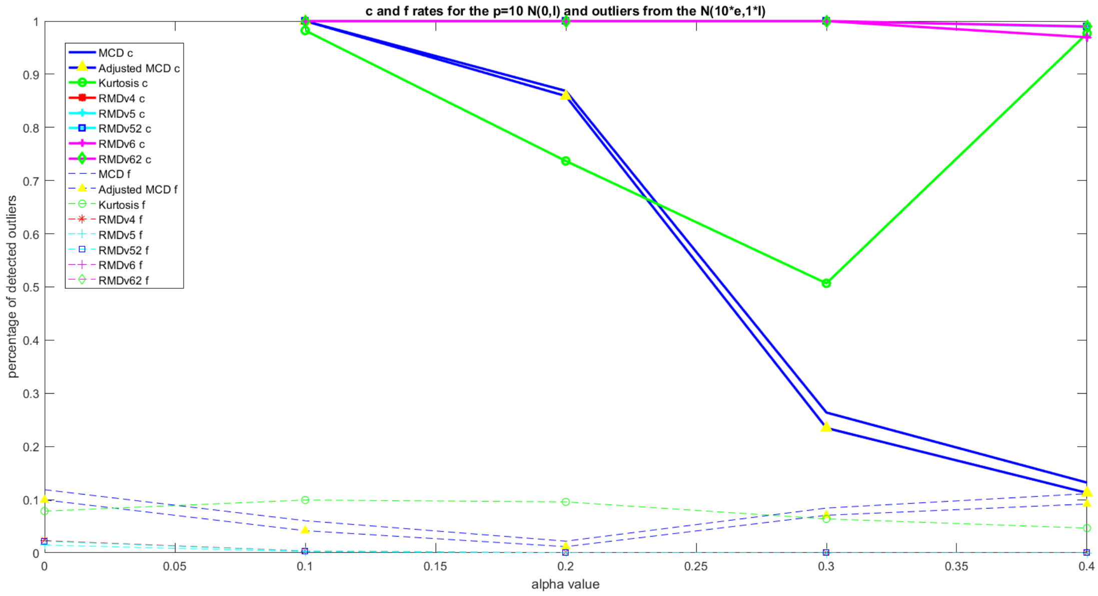

Figure 8: Percentages of detected outliers (c and f) for each method.

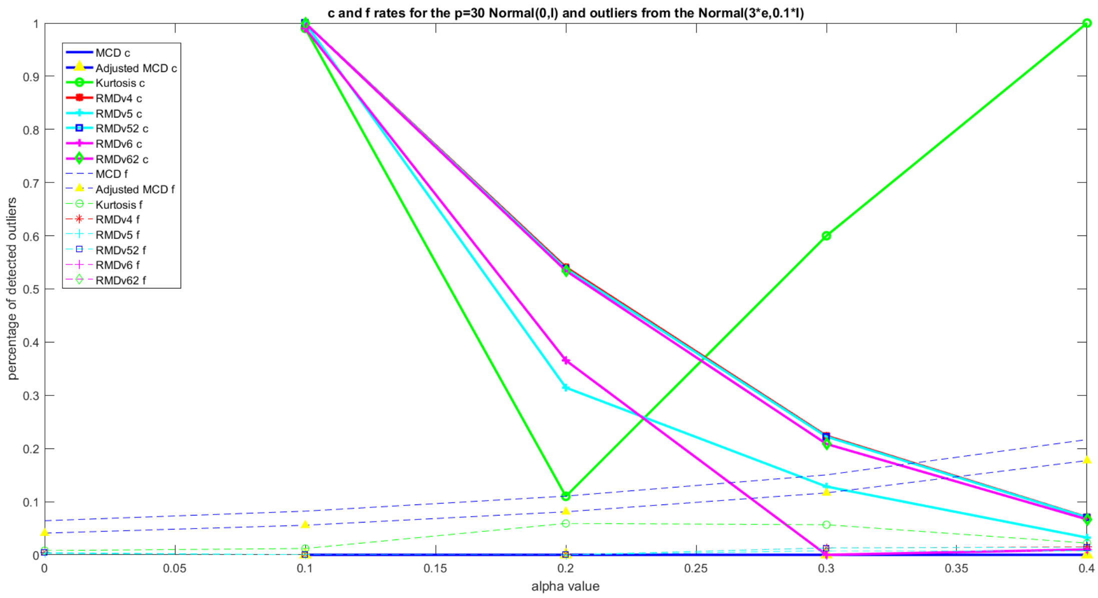

Figure 9: Percentages of detected outliers (c and f) for each method.

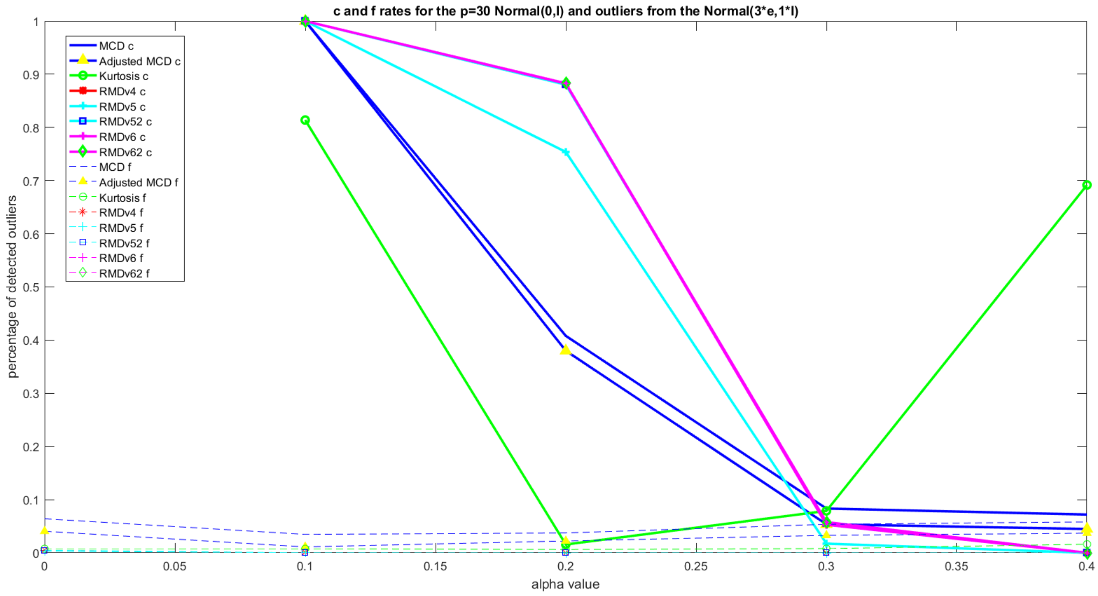

Figure 10: Percentages of detected outliers (c and f) for each method.

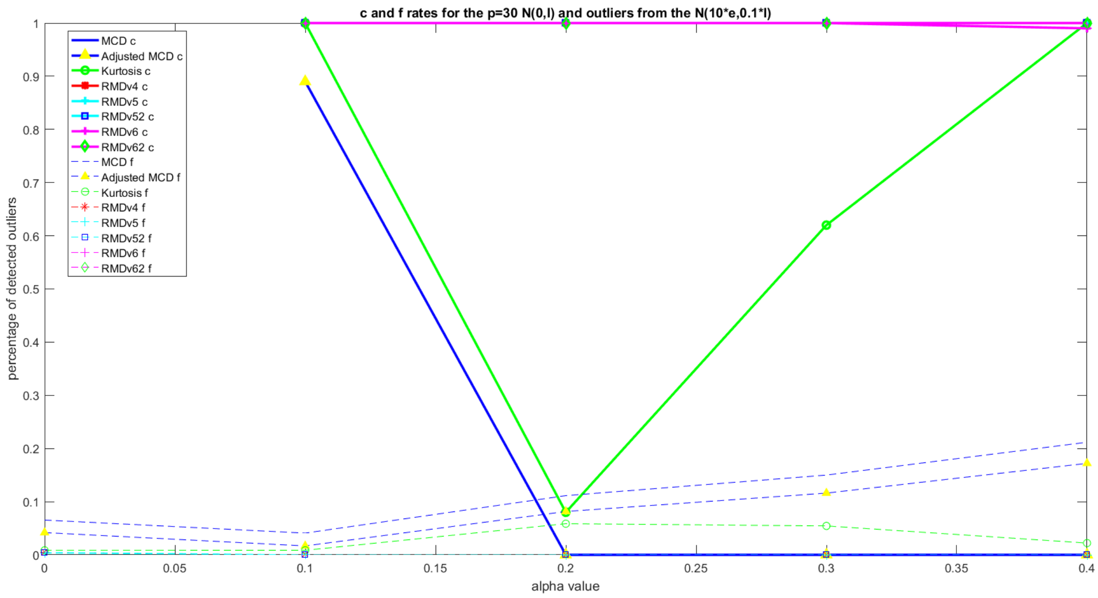

Figure 11: Percentages of detected outliers (c and f) for each method.

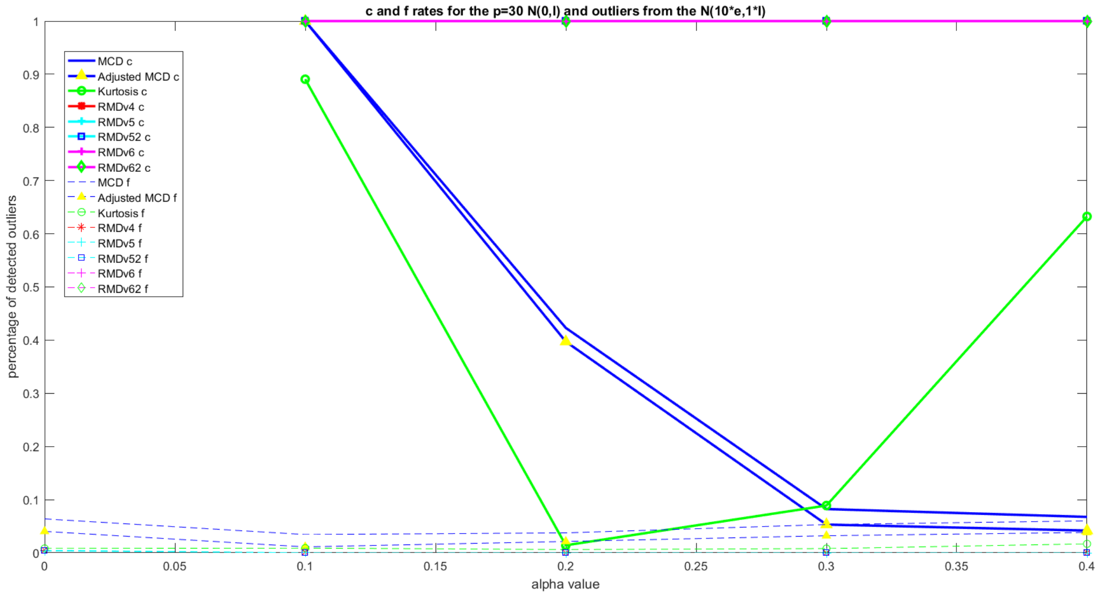

Figure 12: Percentages of detected outliers (c and f) for each method.

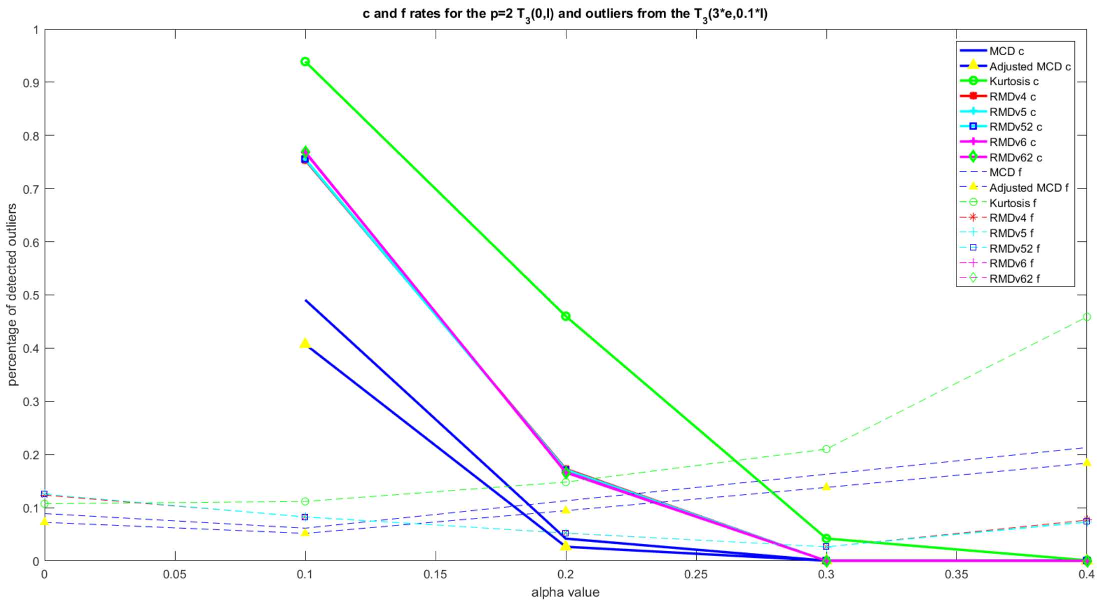

Figure 13: Percentages of detected outliers (c and f) for each method.

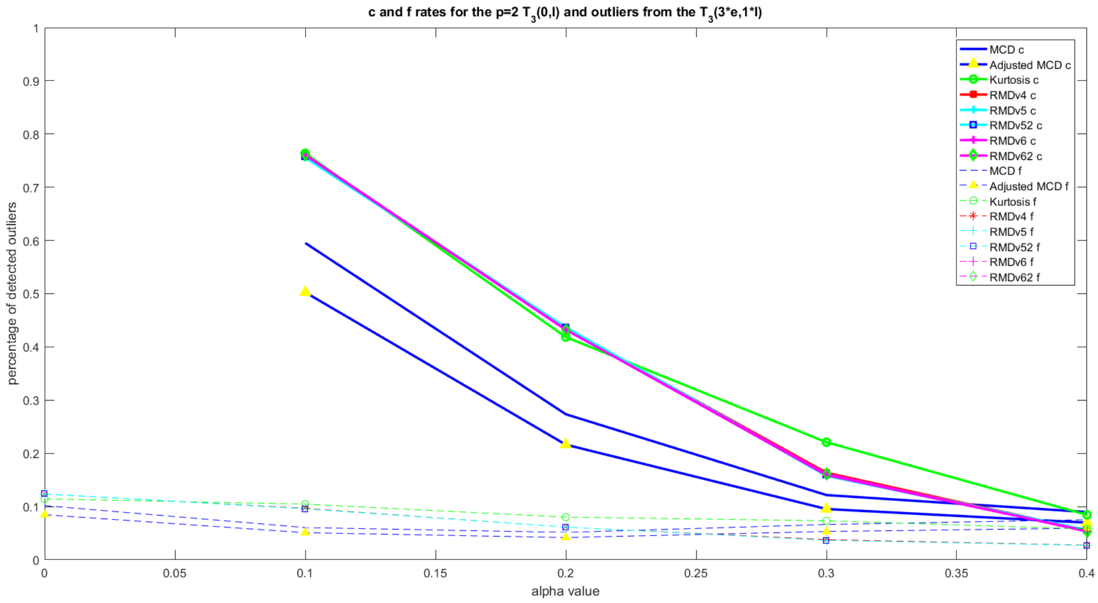

Figure 14: Percentages of detected outliers (c and f) for each method.

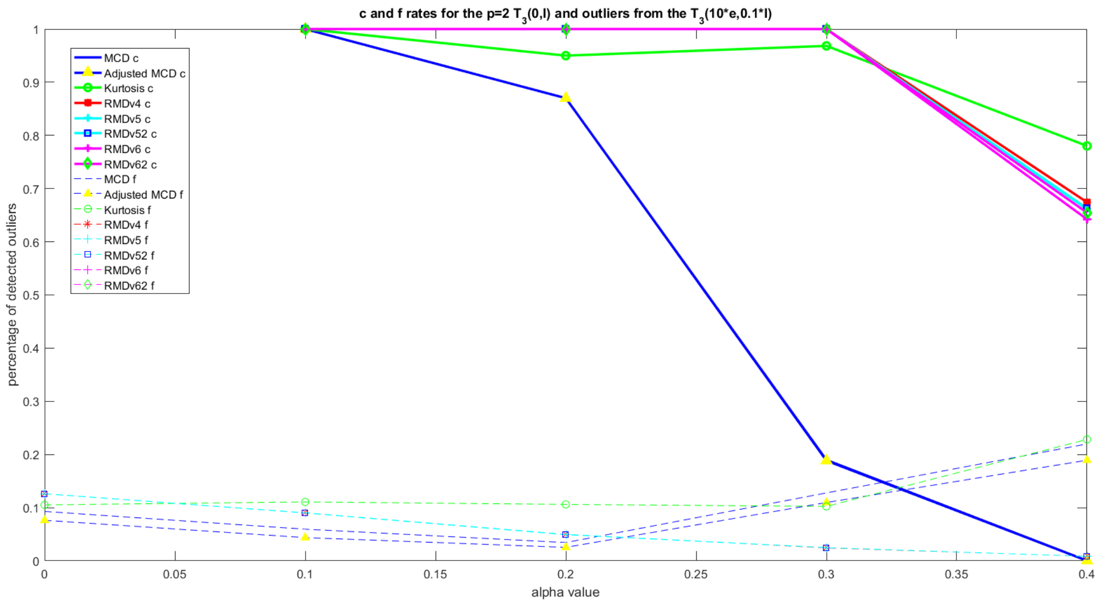

Figure 15: Percentages of detected outliers (c and f) for each method.

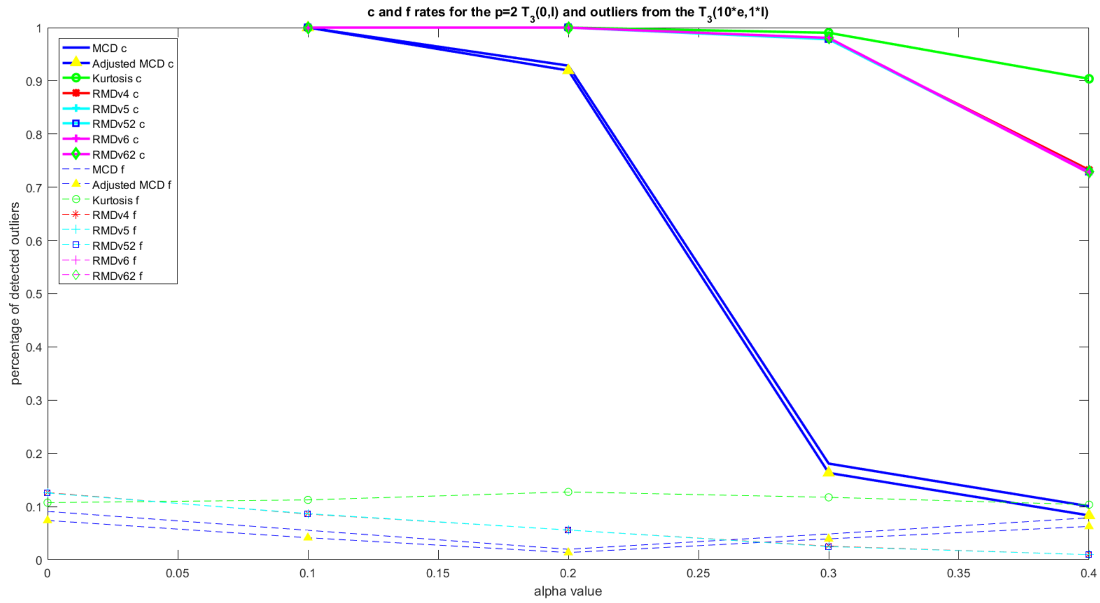

Figure 16: Percentages of detected outliers (c and f) for each method.

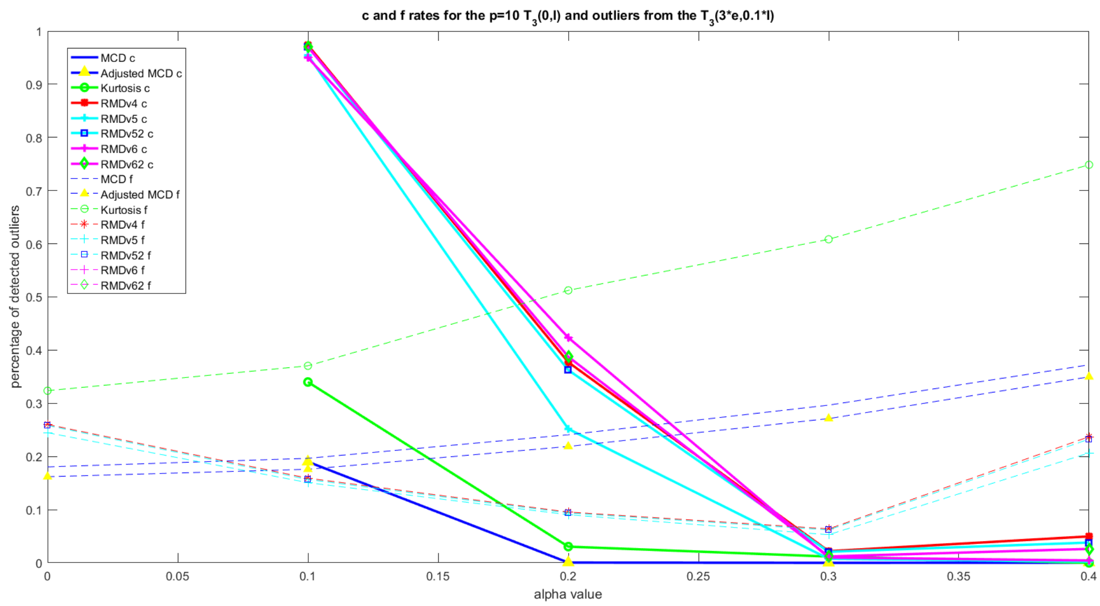

Figure 17: Percentages of detected outliers (c and f) for each method.

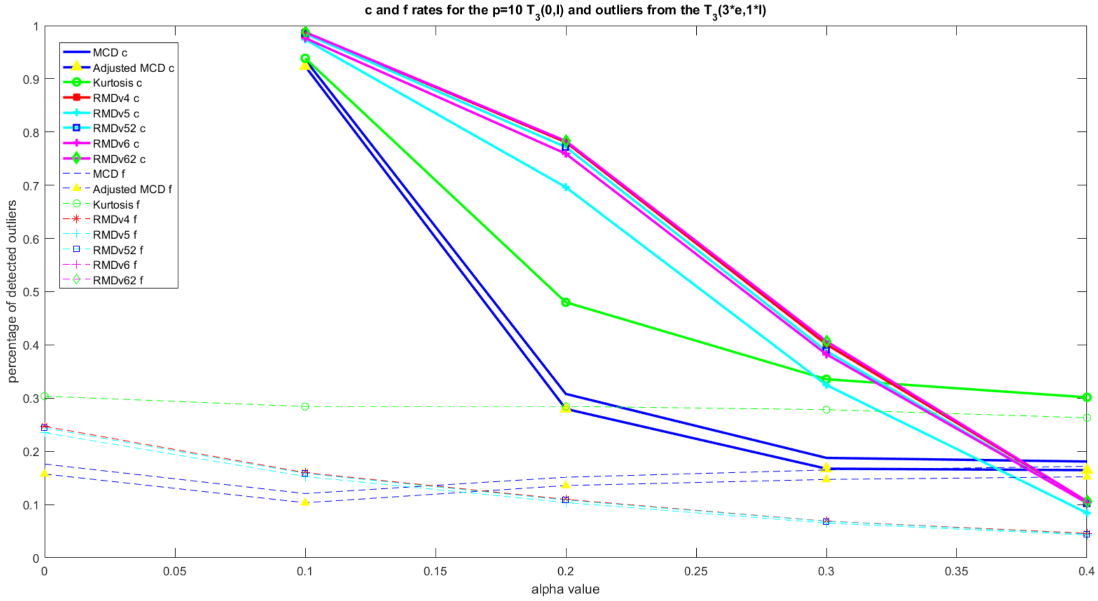

Figure 18: Percentages of detected outliers (c and f) for each method.

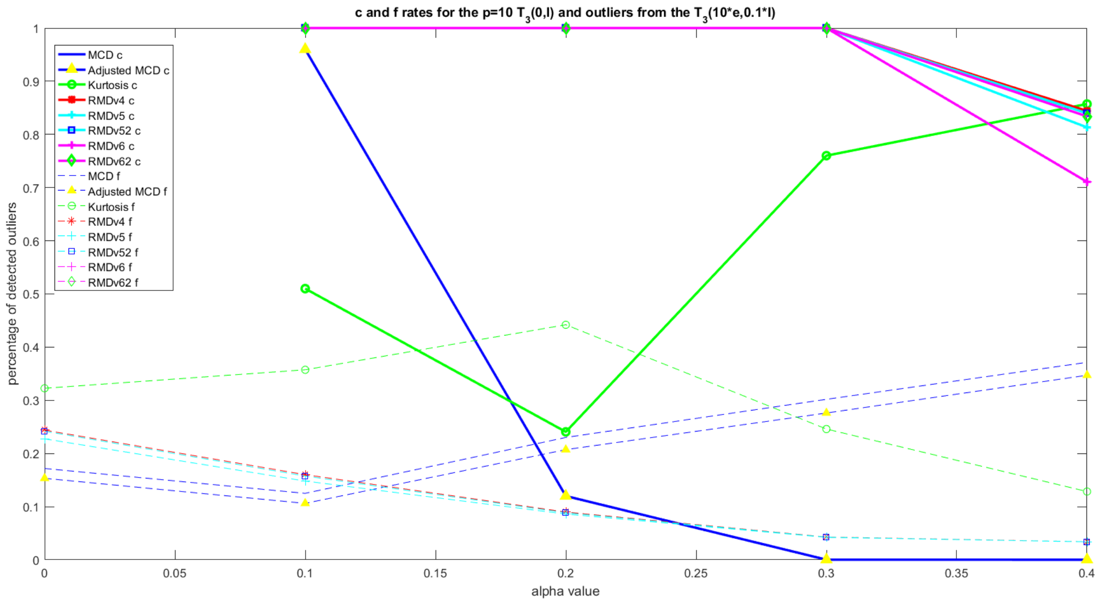

Figure 19: Percentages of detected outliers (c and f) for each method.

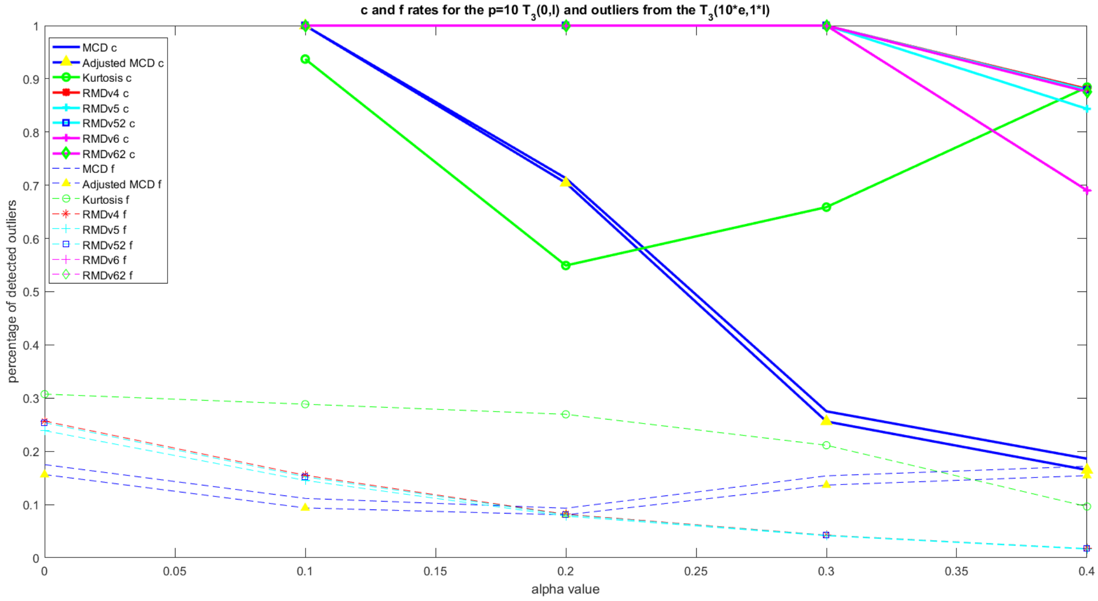

Figure 20: Percentages of detected outliers (c and f) for each method.

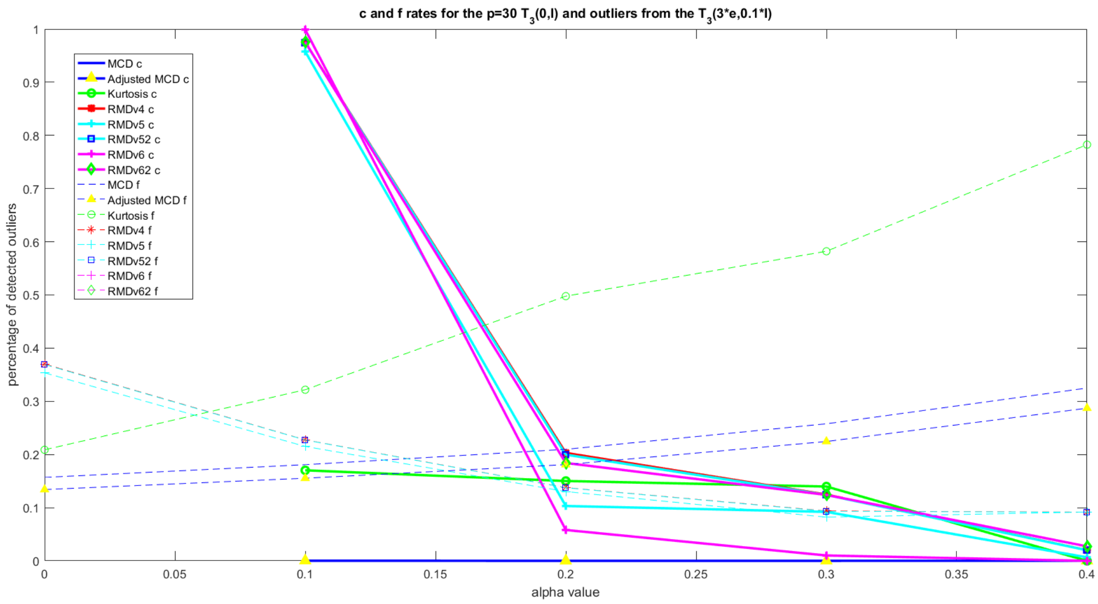

Figure 21: Percentages of detected outliers (c and f) for each method.

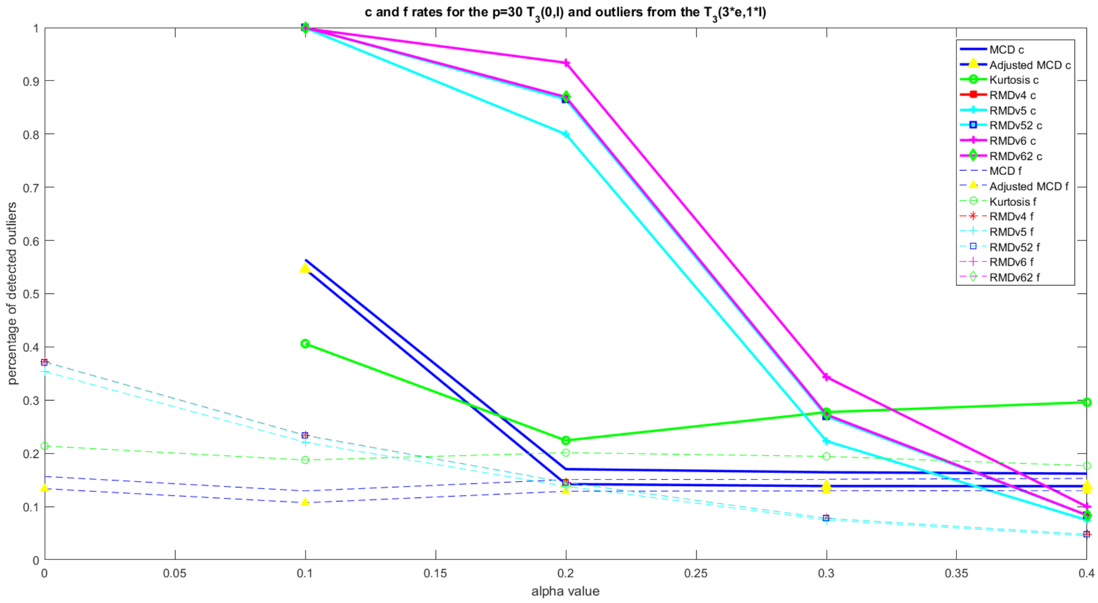

Figure 22: Percentages of detected outliers (c and f) for each method.

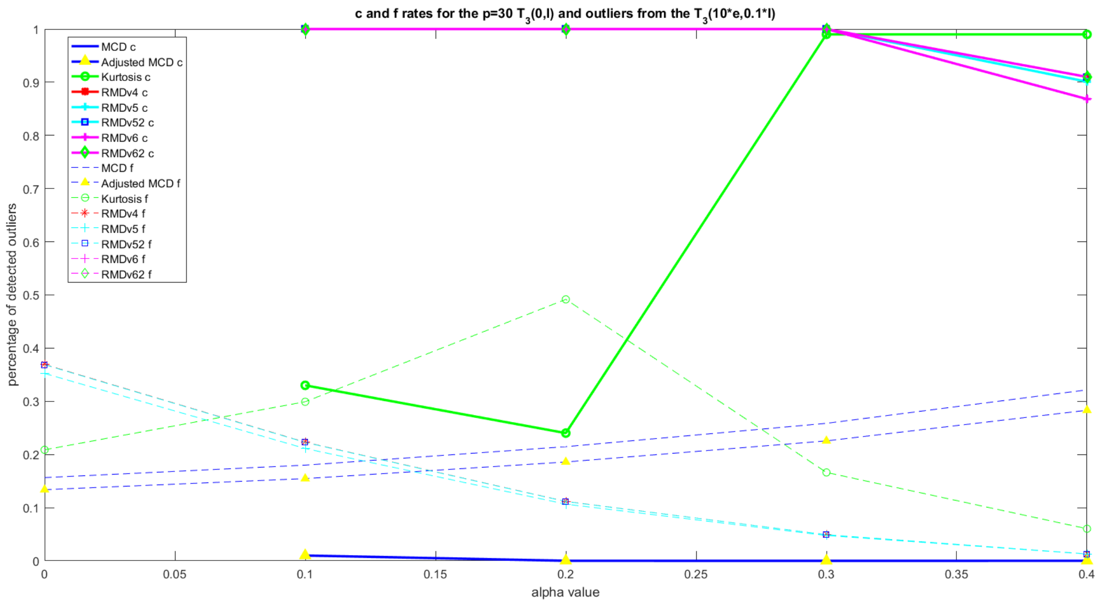

Figure 23: Percentages of detected outliers (c and f) for each method.

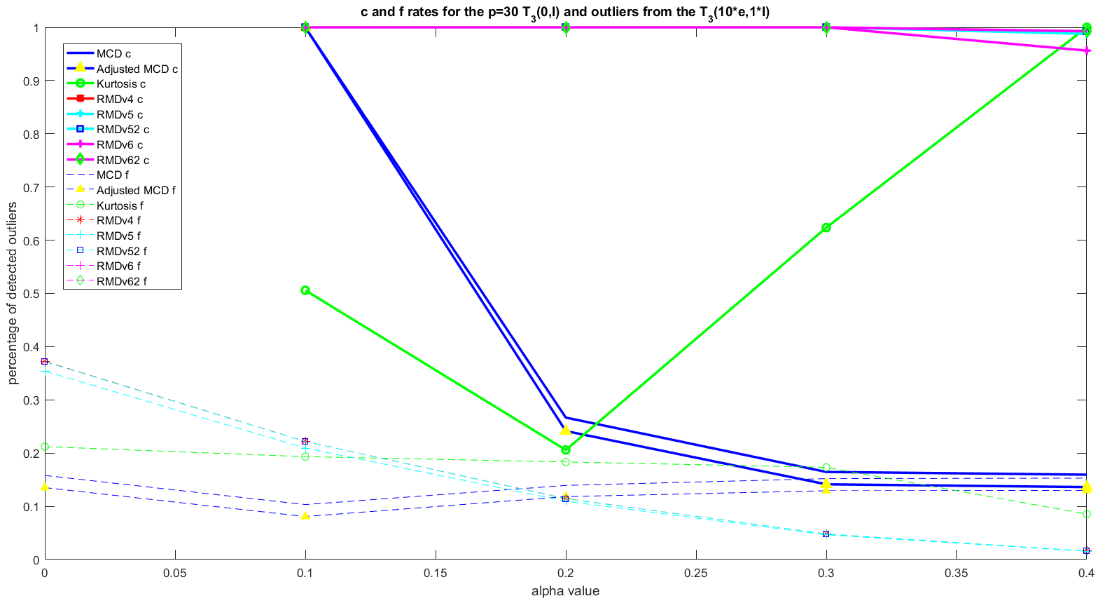

Figure 24: Percentages of detected outliers (c and f) for each method.

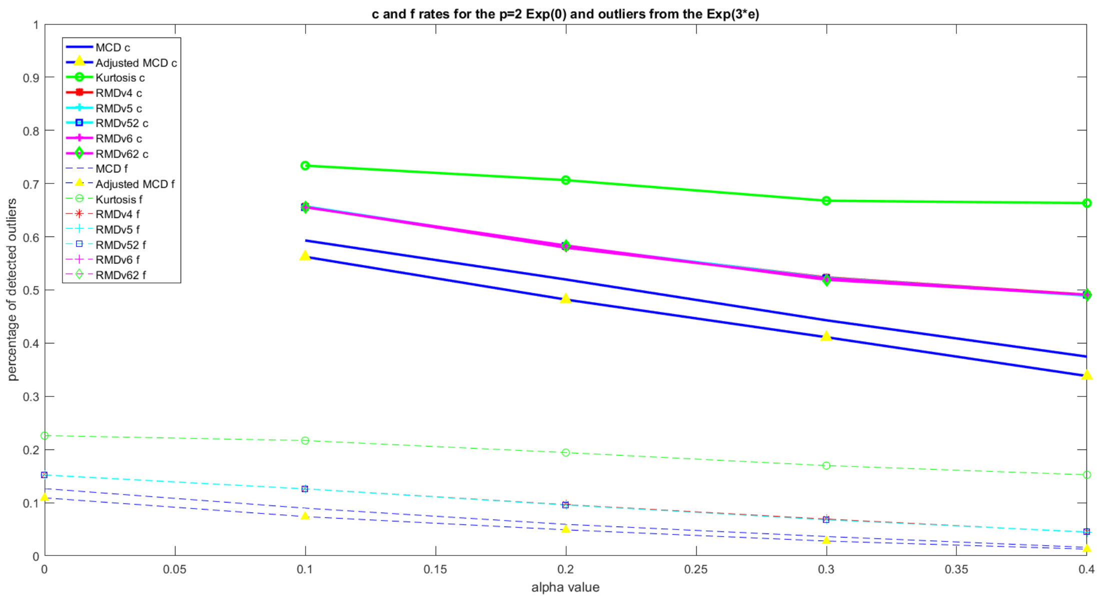

Figure 25: Percentages of detected outliers (c and f) for each method.

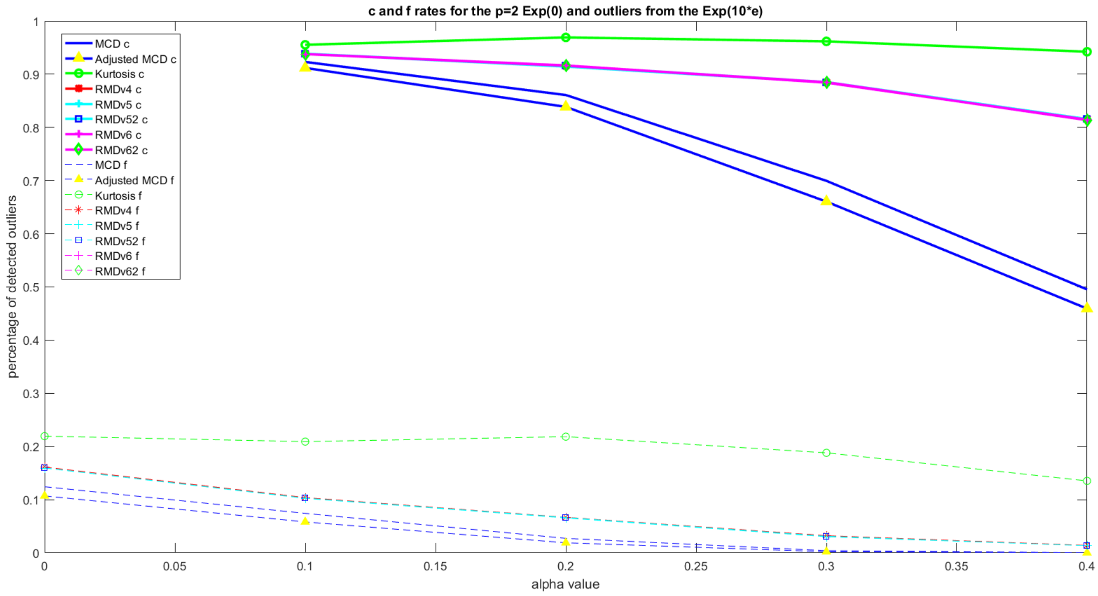

Figure 26: Percentages of detected outliers (c and f) for each method.

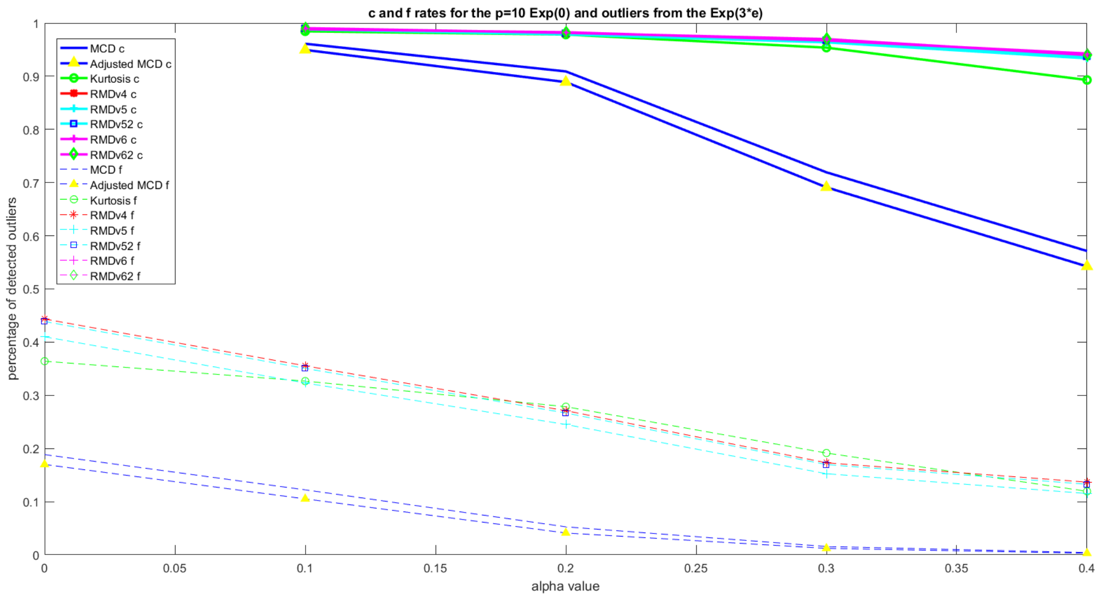

Figure 27: Percentages of detected outliers (c and f) for each method.

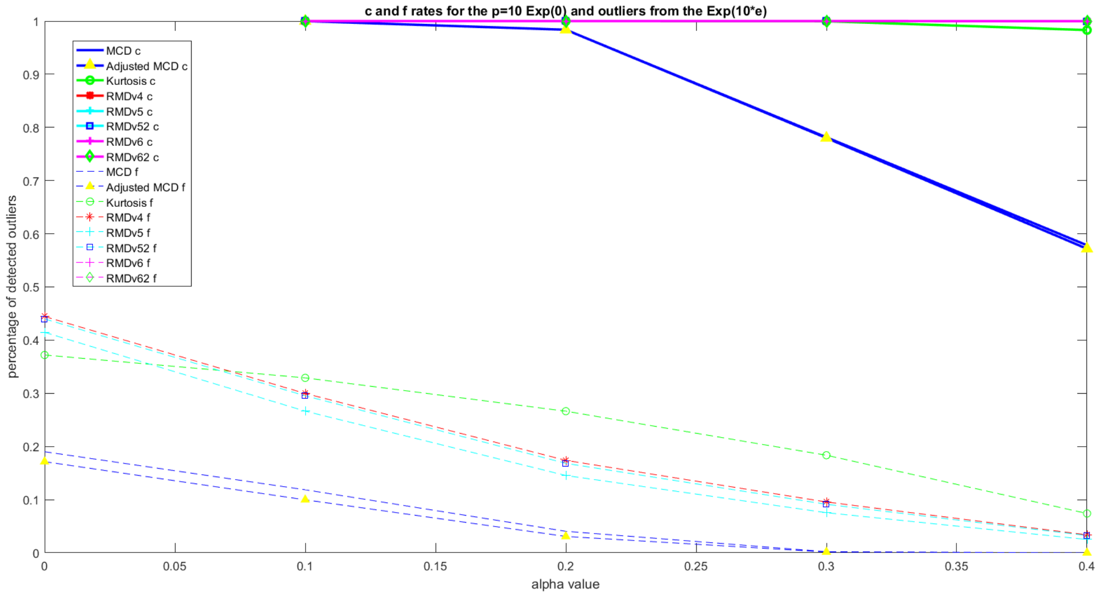

Figure 28: Percentages of detected outliers (c and f) for each method.

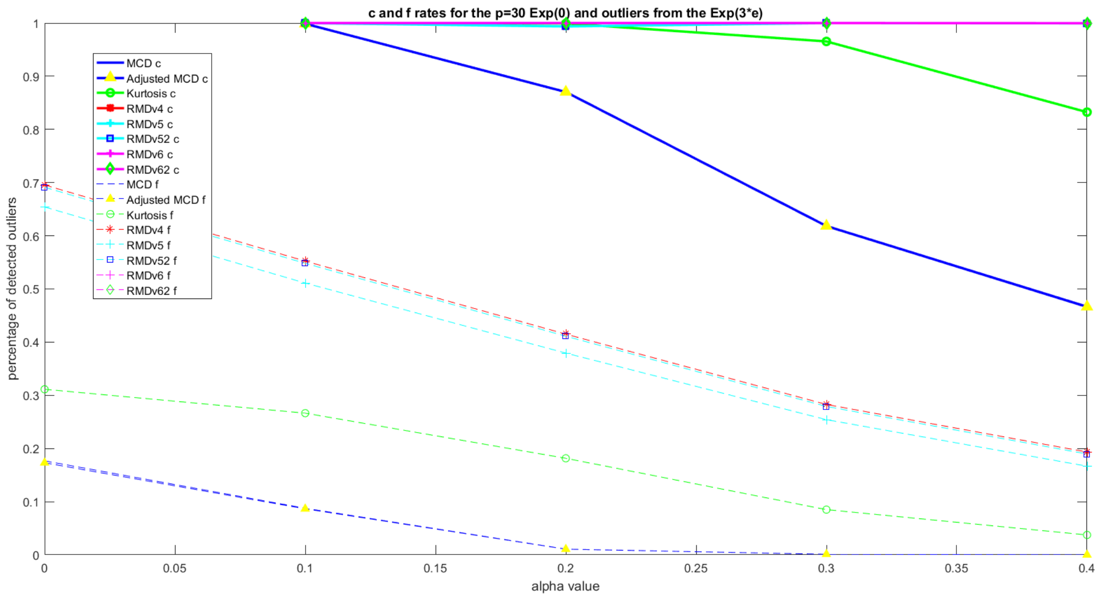

Figure 29: Percentages of detected outliers (c and f) for each method.

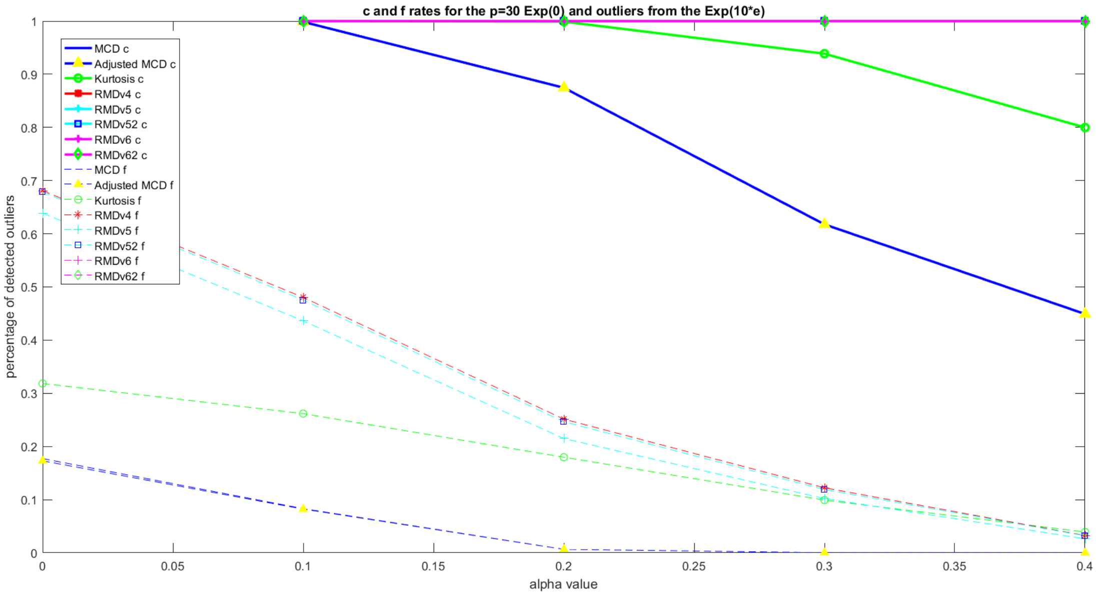

Figure 30: Percentages of detected outliers (c and f) for each method.

Real dataset example

The following figures illustrate the results for the real dataset example from Section 5.

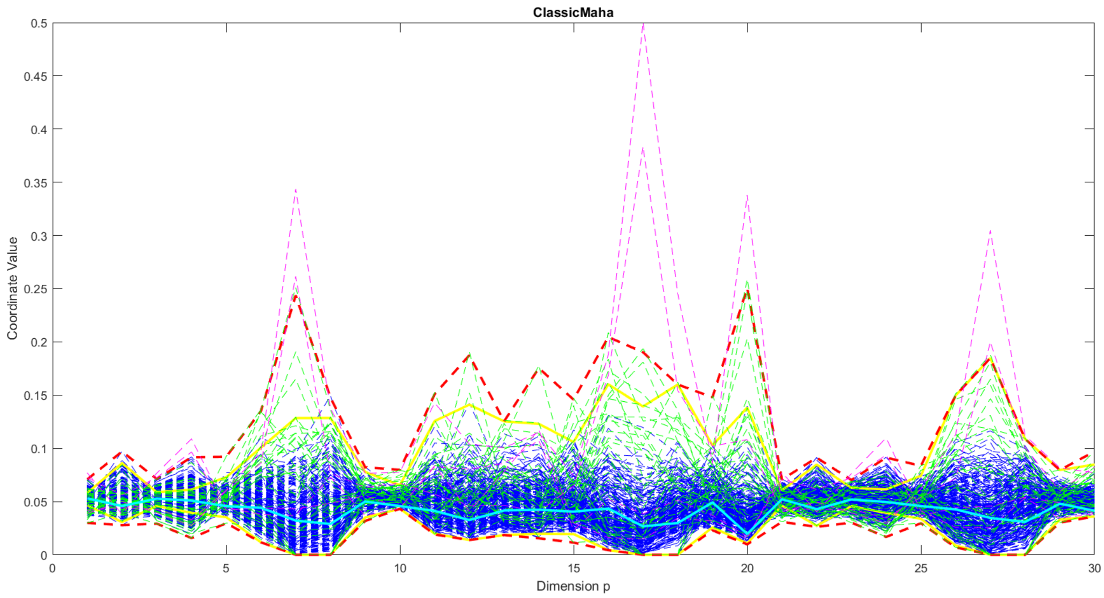

Figure 31: Detected outliers by ClassicMaha.

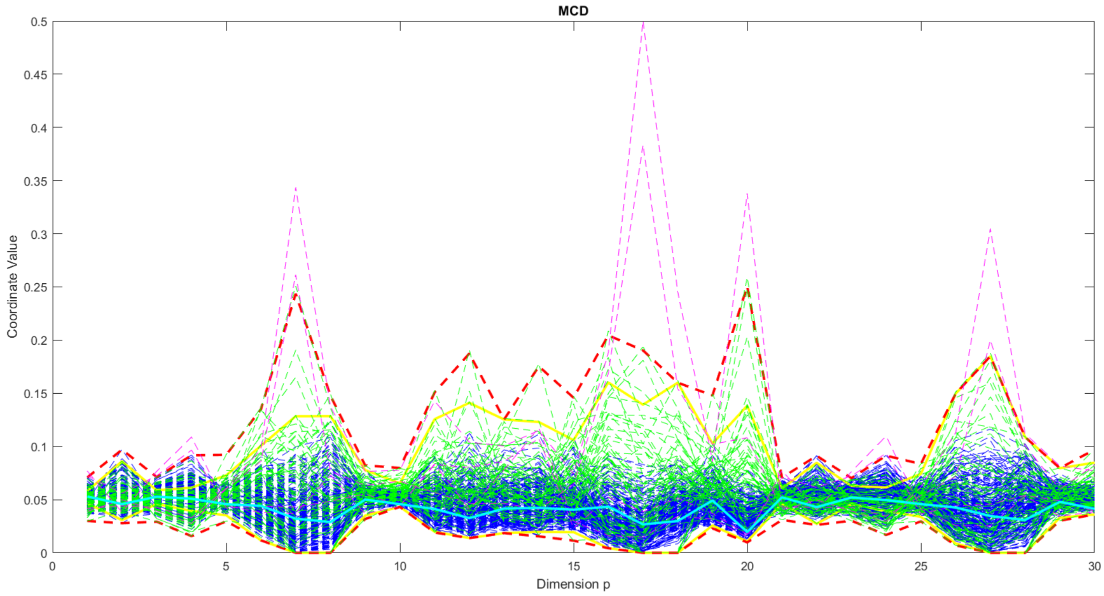

Figure 32: Detected outliers by MCD.

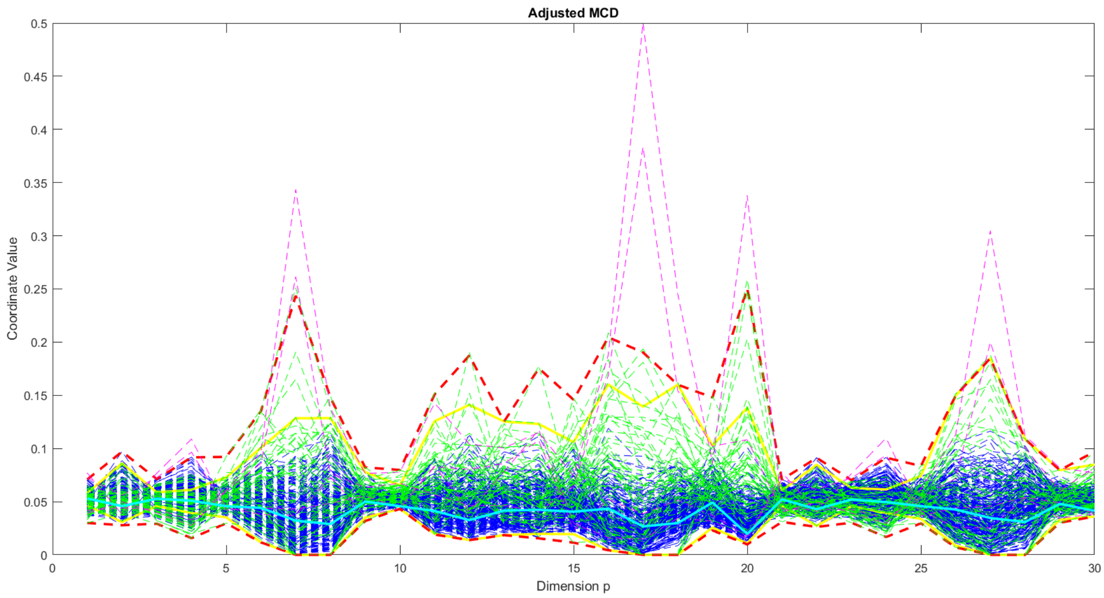

Figure 33: Detected outliers by Adjusted MCD.

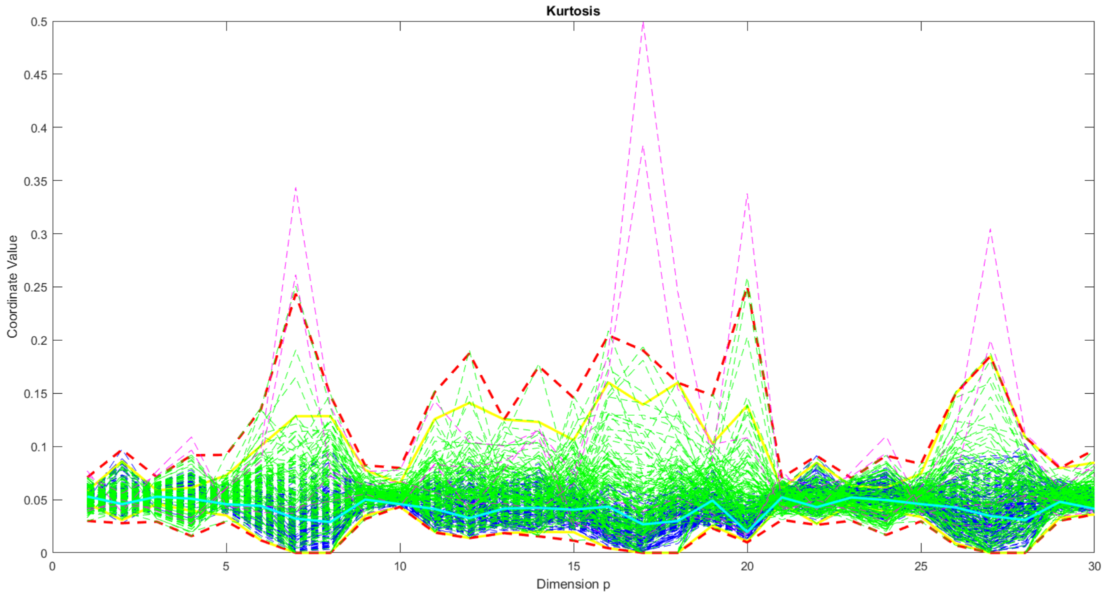

Figure 34: Detected outliers by Kurtosis.

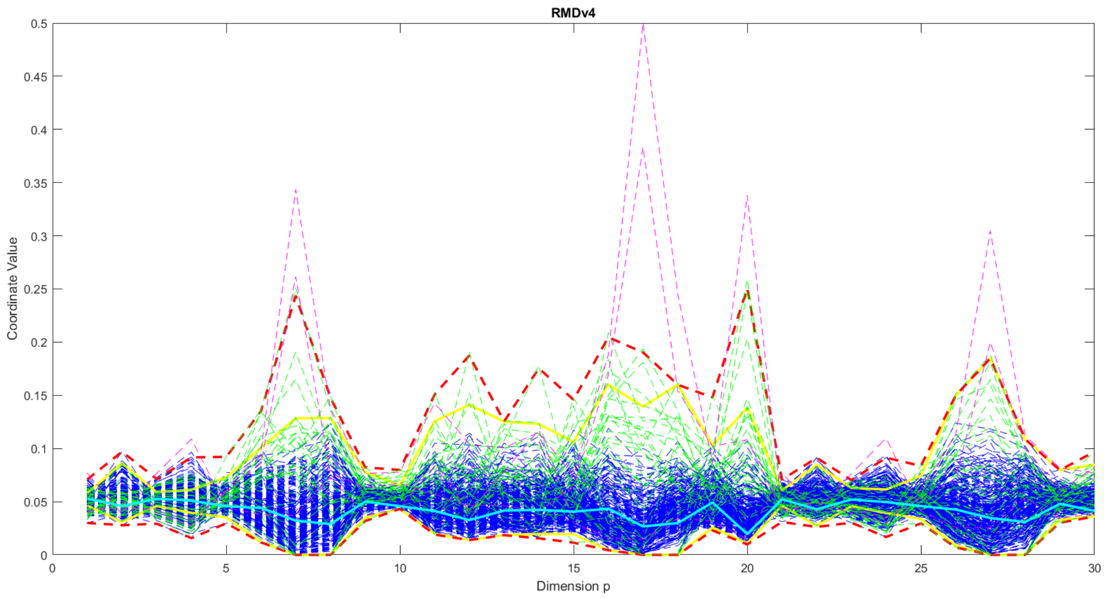

Figure 35: Detected outliers by RMDv4.

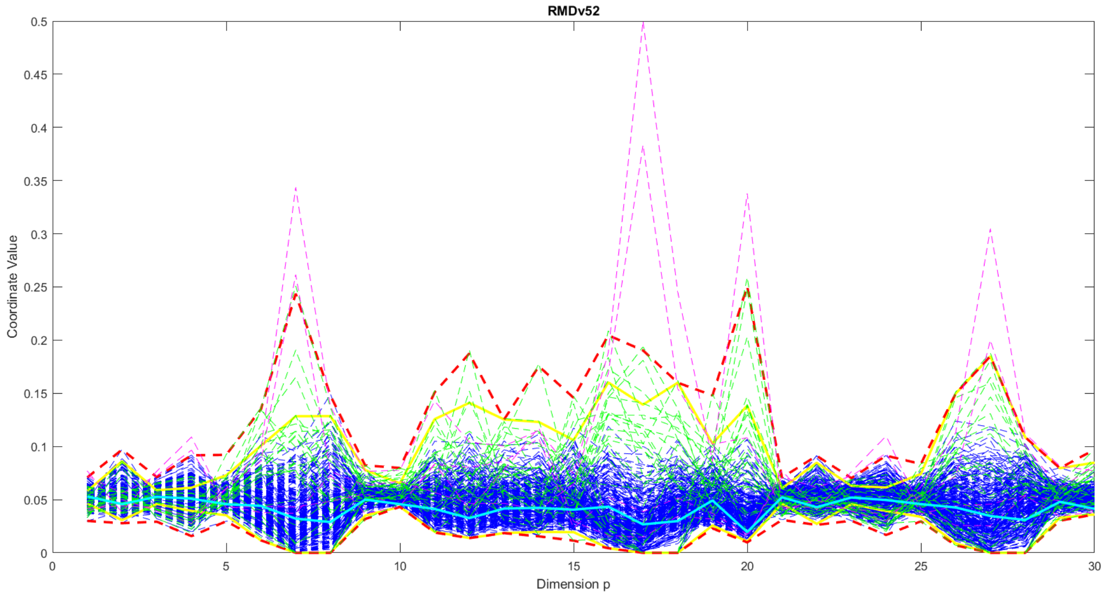

Figure 36: Detected outliers by RMDv52.

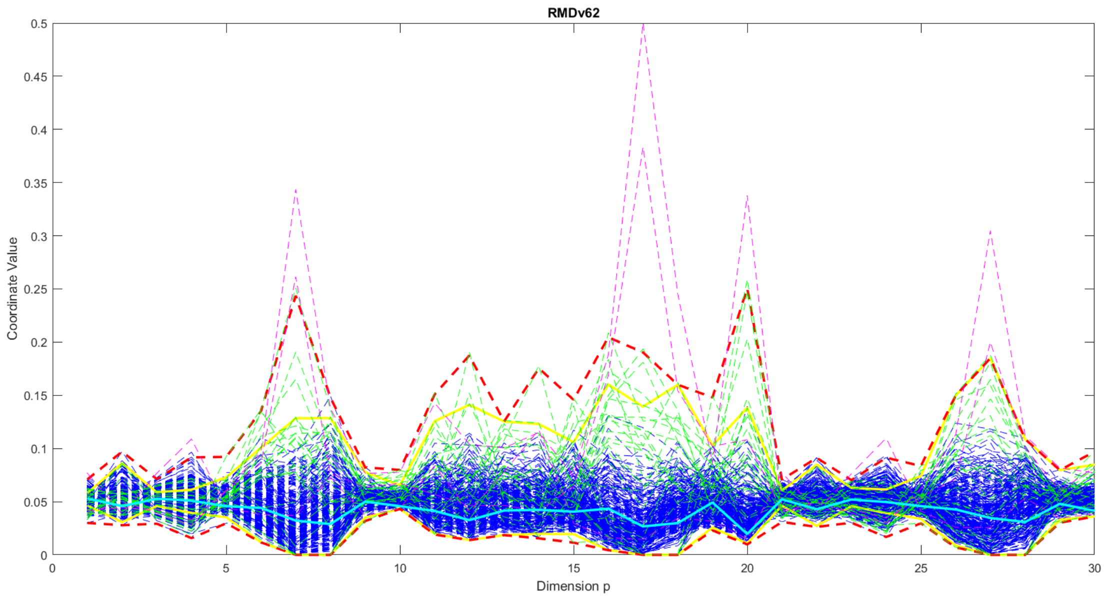

Figure 37: Detected outliers by RMDv62.

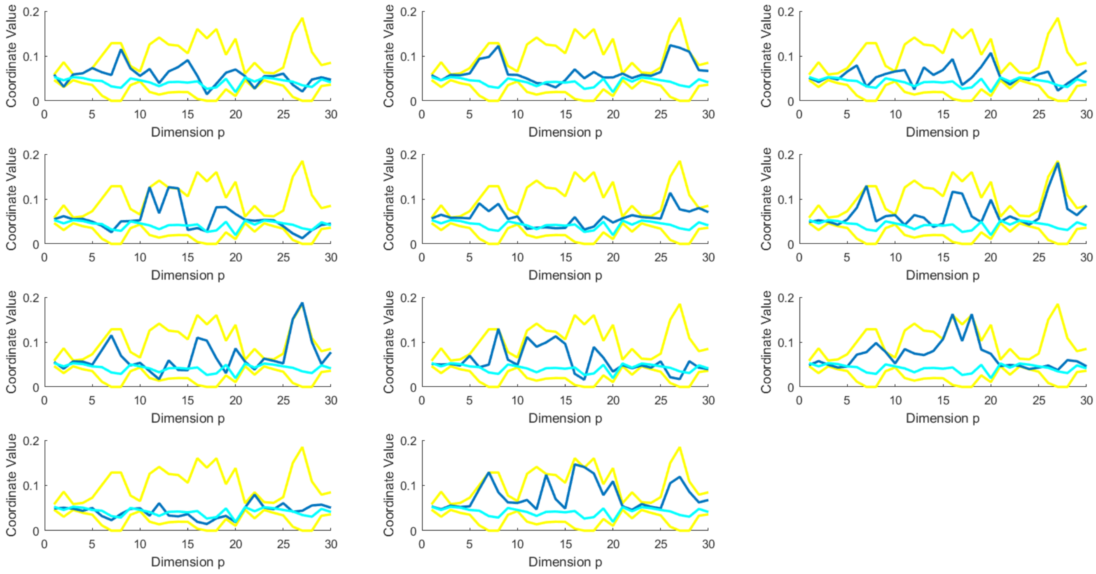

Figure 38: ClassicMaha detected outliers that belong to the 50% of the most central data.

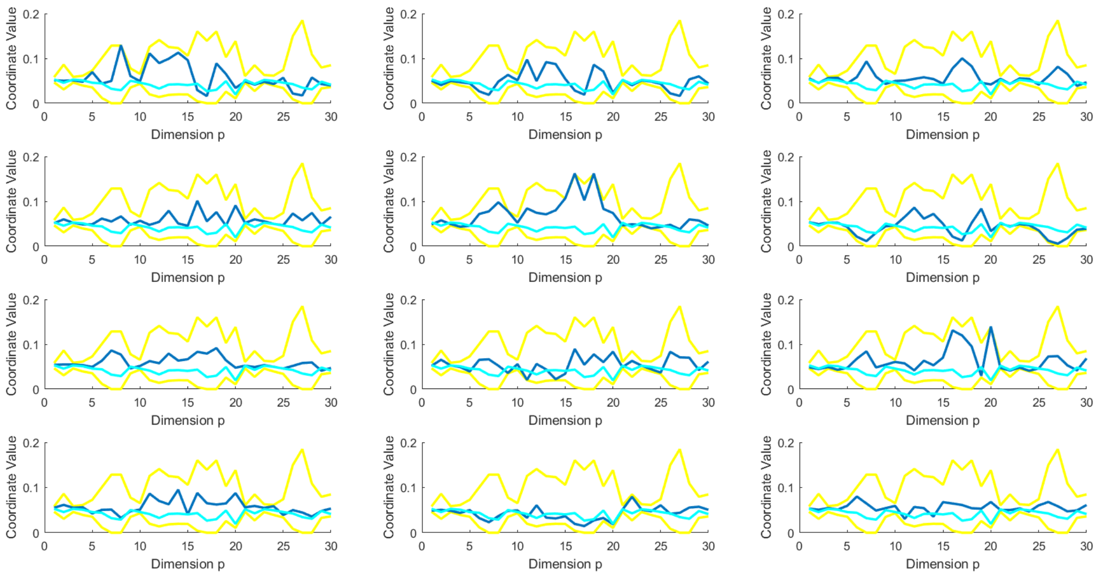

Figure 39: MCD detected outliers that belong to the 50% of the most central data.

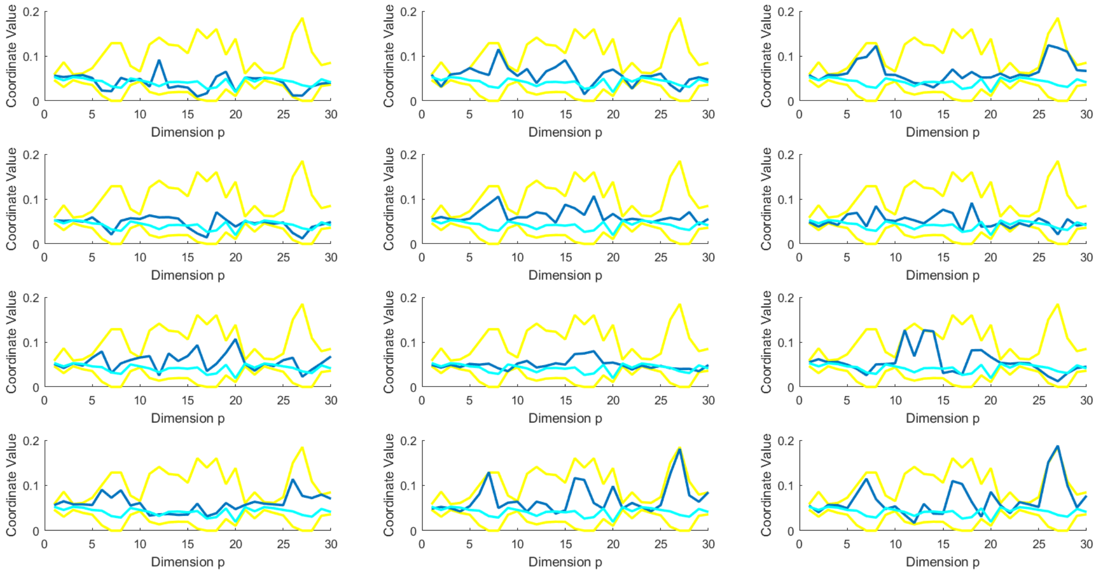

Figure 40: MCD detected outliers that belong to the 50% of the most central data.

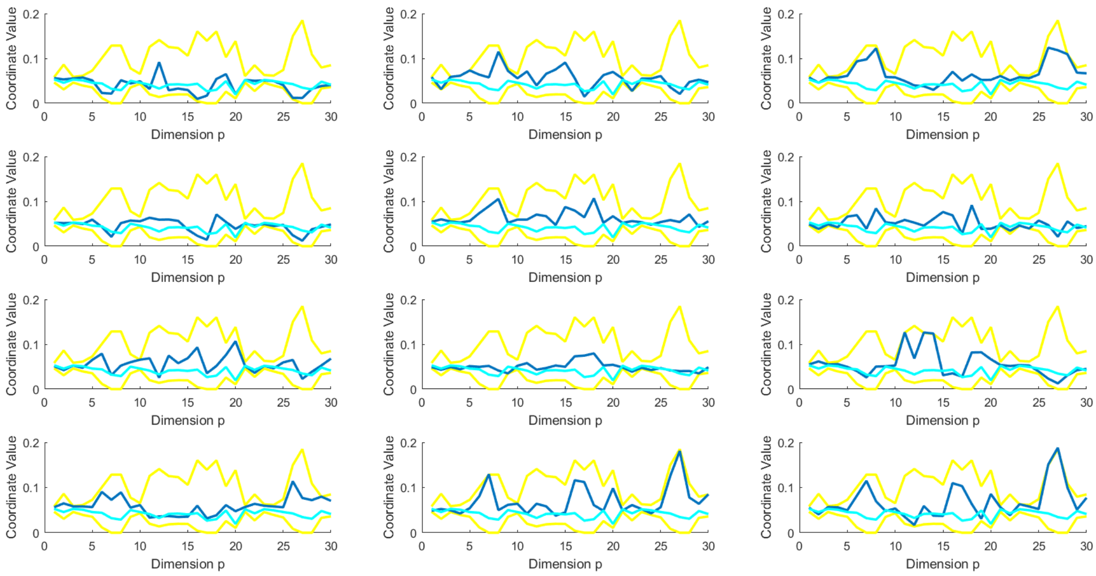

Figure 41: Adjusted MCD detected outliers that belong to the 50% of the most central data.

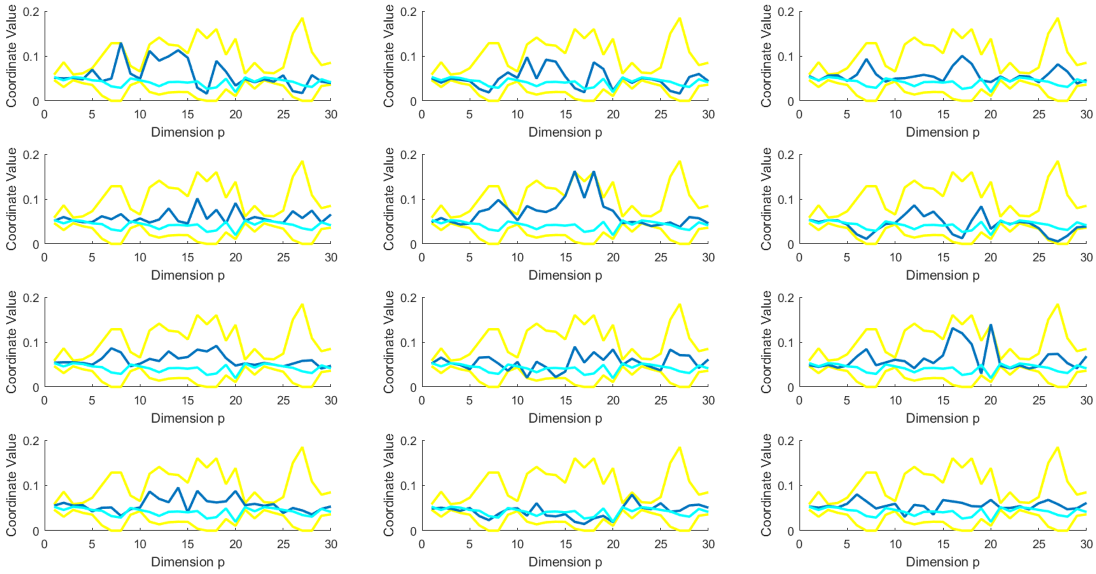

Figure 42: Adjusted MCD detected outliers that belong to the 50% of the most central data.

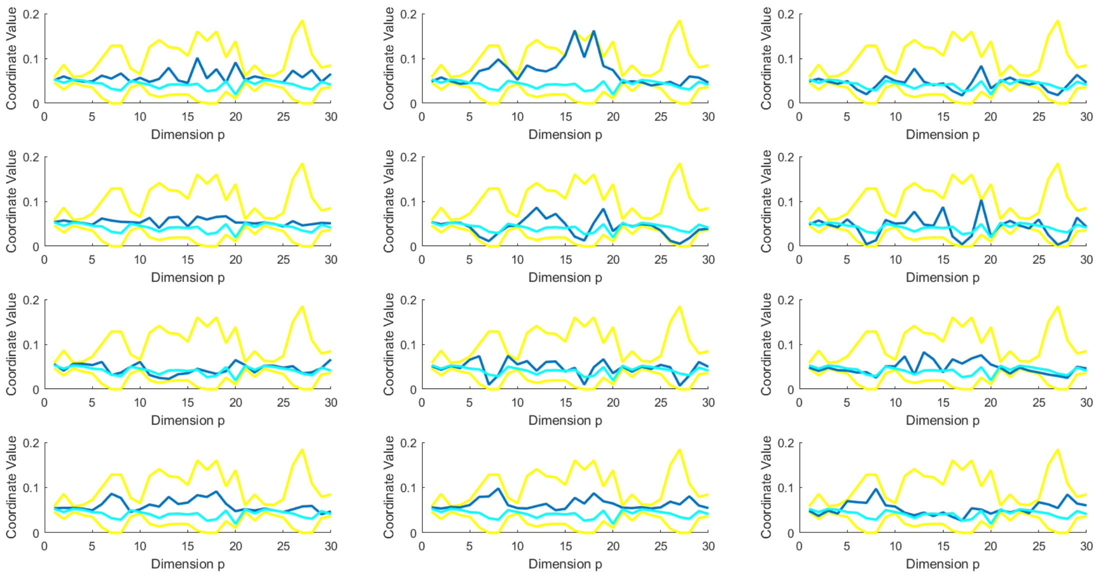

Figure 43: Kurtosis detected outliers that belong to the 50% of the most central data.

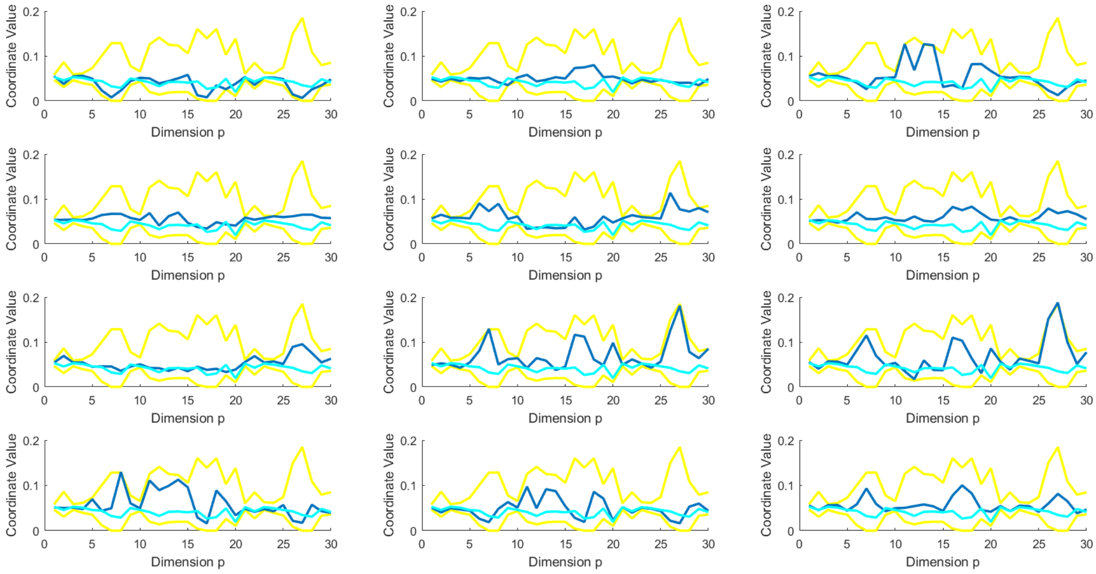

Figure 44: Kurtosis detected outliers that belong to the 50% of the most central data.

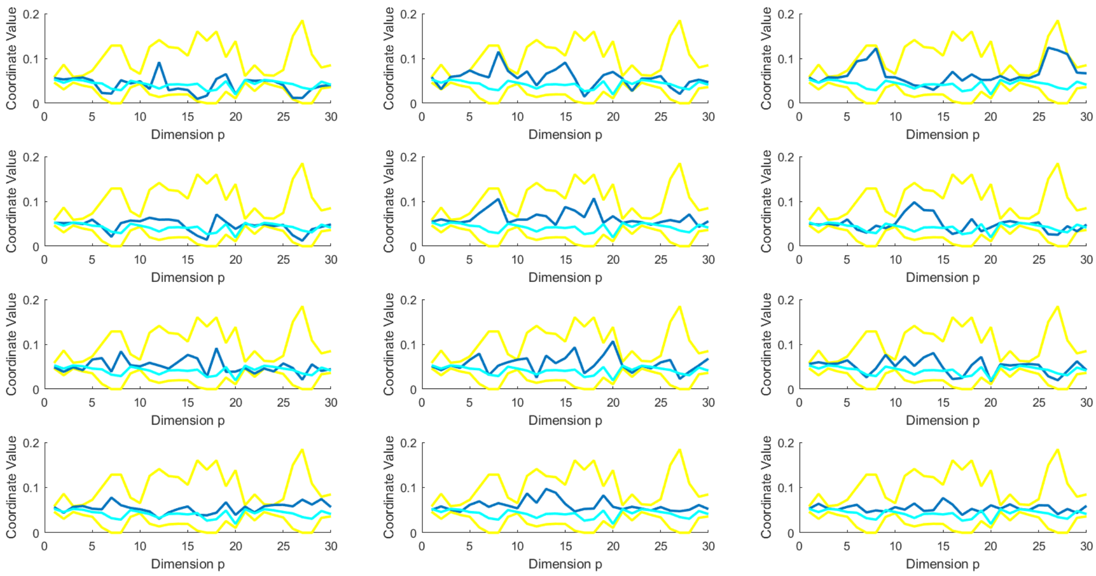

Figure 45: Kurtosis detected outliers that belong to the 50% of the most central data.

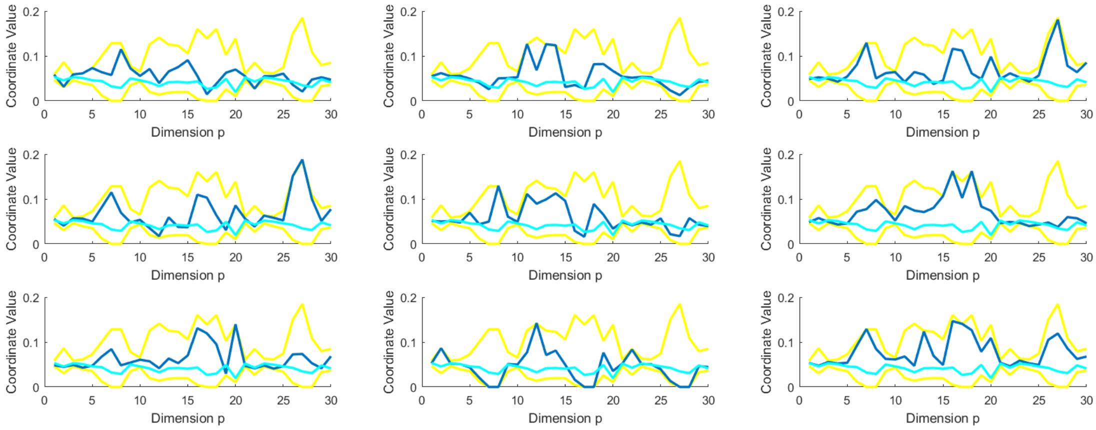

Figure 46: RMDv4 detected outliers that belong to the 50% of the most central data.

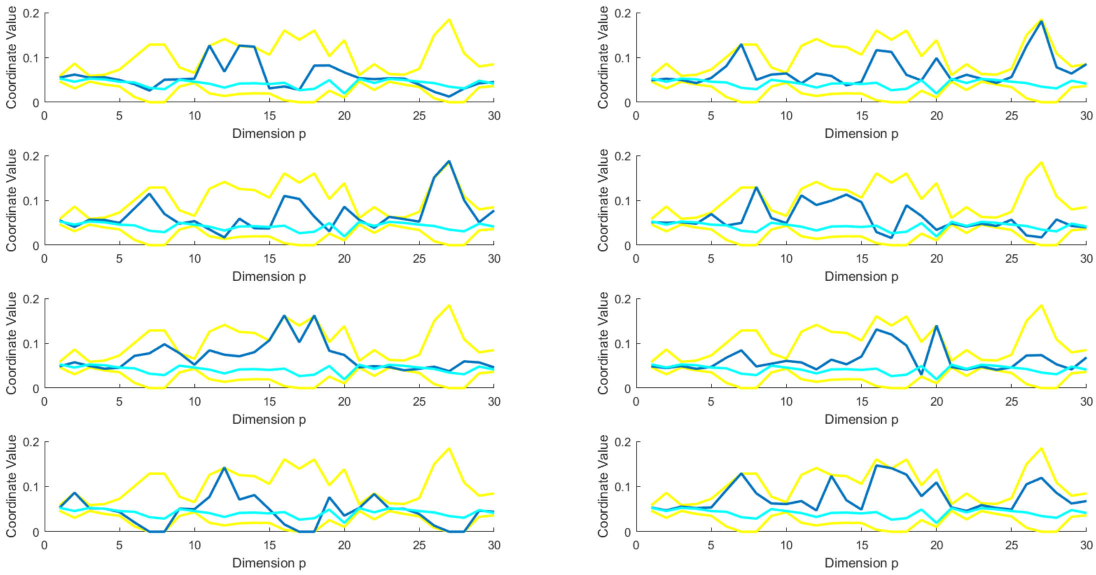

Figure 47: RMDv52 detected outliers that belong to the 50% of the most central data.

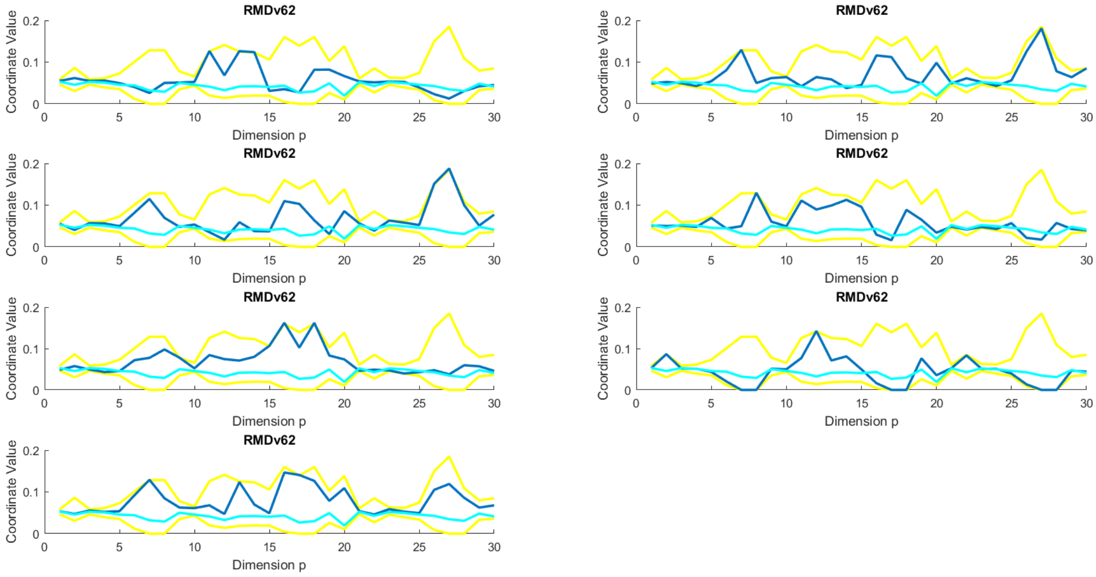

Figure 48: RMDv62 detected outliers that belong to the 50% of the most central data.

References

C. Agostinelli and M. Romanazzi. Local depth. *Journal of Statistical Planning and Inference*, 141(2):817–830, 2011.

S. D. Bay. The uci kdd archive [http://kdd. ics. uci. edu]. irvine, ca: University of california. *Department of Information and Computer Science*, 404:405, 1999.

C. Becker and U. Gather. The masking breakdown point of multivariate outlier identification rules. *Journal of the American Statistical Association*, 94(447): 947–955, 1999.

A. Bose. Estimating the asymptotic dispersion of the l1 median. *Annals of the Institute of Statistical Mathematics*, 47(2):267–271, 1995.

A. Bose and P. Chaudhuri. On the dispersion of multivariate median. *Annals of the Institute of Statistical Mathematics*, 45(3):541–550, 1993.

J. Brettschneider, F. Collin, B. M. Bolstad, and T. P. Speed. Quality assessment for short oligonucleotide microarray data. *Technometrics*, 50(3):241–264, 2008.

B. Brown. Statistical uses of the spatial median. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 25–30, 1983.

E. A. Cator and H. P. Lopuhaä. Asymptotic expansion of the minimum covariance determinant estimators. *Journal of Multivariate Analysis*, 101(10):2372–2388, 2010.

E. A. Cator, H. P. Lopuhaä, et al. Central limit theorem and influence function for the mcd estimators at general multivariate distributions. *Bernoulli*, 18(2): 520–551, 2012.

A. Cerioli, M. Riani, A. C. Atkinson, D. Perrotta, and F. Torti. Fitting mixtures of regression lines with the forward search. *Mining Massive Data Sets for Security: Advances in Data Mining, Search, Social Networks and Text Mining, and Their Applications to Security*, 19:271, 2008.

A. Cerioli, M. Riani, and A. C. Atkinson. Controlling the size of multivariate outlier tests with the mcd estimator of scatter. *Statistics and Computing*, 19(3): 341–353, 2009.

S. X. Chen, Y.-L. Qin, et al. A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics*, 38(2):808–835, 2010.

Y. Chen, X. Dang, H. Peng, and H. L. Bart. Outlier detection with the kernelized spatial depth function. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):288–305, 2009.

H. C. Choi, H. P. Edwards, C. H. Sweatman, and V. Obolonkin. Multivariate outlier detection of dairy herd testing data. *ANZIAM Journal*, 57:38–53, 2016.

J. T. Chu. On the distribution of the sample median. *The Annals of Mathematical Statistics*, pages 112–116, 1955.

V. DeMiguel, A. Martin-Utrera, and F. J. Nogales. Size matters: Optimal calibration of shrinkage estimators for portfolio selection. *Journal of Banking & Finance*, 37(8):3018–3034, 2013.

Y. Dodge. An introduction to l1-norm based statistical data analysis. *Computational Statistics & Data Analysis*, 5(4):239–253, 1987.

M. Falk. On mad and comedians. *Annals of the Institute of Statistical Mathematics*, 49(4):615–644, 1997.

P. Filzmoser, R. G. Garrett, and C. Reimann. Multivariate outlier detection in exploration geochemistry. *Computers & geosciences*, 31(5):579–587, 2005.

C. F. Gauss. Bestimmung der genauigkeit der beobachtungen. *Astronomi*, 1: 185–197, 1816.

R. Gnanadesikan and J. R. Kettenring. Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics*, pages 81–124, 1972.

C. Goutte and E. Gaussier. A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In *European Conference on Information Retrieval*, pages 345–359. Springer, 2005.

J. Gower. Algorithm as 78: The mediancentre. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 23(3):466–470, 1974.

A. S. Hadi. Identifying multiple outliers in multivariate data. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 761–771, 1992.

P. Hall and A. Welsh. Limit theorems for the median deviation. *Annals of the Institute of Statistical Mathematics*, 37(1):27–36, 1985.

J. Hardin and D. M. Rocke. The distribution of robust distances. *Journal of Computational and Graphical Statistics*, 14(4):928–946, 2005.

M. Hubert and M. Debruyne. Minimum covariance determinant. *Wiley interdisciplinary reviews: Computational statistics*, 2(1):36–43, 2010.

M. Hubert, P. J. Rousseeuw, and S. Van Aelst. High-breakdown robust multivariate methods. *Statistical Science*, pages 92–119, 2008.

A. Inselberg. *Parallel coordinates*. Springer, 2009.

A. Inselberg and B. Dimsdale. Parallel coordinates: a tool for visualizing multidimensional geometry. In *Proceedings of the 1st conference on Visualization'90*, pages 361–378. IEEE Computer Society Press, 1990.

W. James and C. Stein. Estimation with quadratic loss. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 361–379, 1961.

N. Lazar. *The statistical analysis of functional MRI data*. Springer Science & Business Media, 2008.

O. Ledoit and M. Wolf. Honey, i shrunk the sample covariance matrix. *UPF economics and business working paper*, (691), 2003a.

O. Ledoit and M. Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance*, 10(5):603–621, 2003b.

O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2):365–411, 2004.

C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4):764–766, 2013.

M. A. Lindquist. The statistical analysis of fmri data. *Statistical Science*, pages 439–464, 2008.

R. Y. Liu et al. On a notion of data depth based on random simplices. *The Annals of Statistics*, 18(1):405–414, 1990.

H. P. Lopuhaa and P. J. Rousseeuw. Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *The Annals of Statistics*, pages 229–248, 1991.

P. C. Mahalanobis. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (Calcutta)*, 2:49–55, 1936.

L. Marcano and W. Fermín. Comparación de métodos de detección de datos anómalos multivariantes mediante un estudio de simulación. *SABER. Revista Multidisciplinaria del Consejo de Investigación de la Universidad de Oriente*, 25 (2):192–201, 2013.

R. A. Maronna and V. J. Yohai. Robust estimation of multivariate location and scatter. *Wiley StatsRef: Statistics Reference Online*, 1976.

R. A. Maronna and R. H. Zamar. Robust estimates of location and dispersion for high-dimensional datasets. *Technometrics*, 44(4):307–317, 2002.

M. M. Monti. Statistical analysis of fmri time-series: a critical review of the glm approach. *Frontiers in human neuroscience*, 5(28), 2011.

J. Möttönen, K. Nordhausen, H. Oja, et al. Asymptotic theory of the spatial median. In *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jurečková*, pages 182–193. Institute of Mathematical Statistics, 2010.

H. Oja. *Multivariate nonparametric methods with R: an approach based on spatial signs and ranks*. Springer Science & Business Media, 2010.

D. Paindaveine and G. Van Bever. From depth to local depth: a focus on centrality. *Journal of the American Statistical Association*, 108(503):1105–1119, 2013.

D. Peña and F. J. Prieto. Multivariate outlier detection and robust covariance matrix estimation. *Technometrics*, 43(3):286–310, 2001.

D. Peña and F. J. Prieto. Combining random and specific directions for outlier detection and robust estimation in high-dimensional multivariate data. *Journal of Computational and Graphical Statistics*, 16(1):228–254, 2007.

D. Perrotta and F. Torti. Detecting price outliers in european trade data with the forward search. In *Data Analysis and Classification*, pages 415–423. Springer, 2010.

J.-B. Poline and M. Brett. The general linear model and fmri: does love last forever? *Neuroimage*, 62(2):871–880, 2012.

D. M. Powers. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. 2011.

C. Reimann and P. Filzmoser. Normal and lognormal data distribution in geochemistry: death of a myth. consequences for the statistical treatment of geochemical and environmental data. *Environmental geology*, 39(9):1001–1014, 2000.

C. Reimann, P. Filzmoser, and R. G. Garrett. Background and threshold: critical comparison of methods of determination. *Science of the Total Environment*, 346 (1):1–16, 2005.

P. Rosseuw and B. Van Zomeren. Unmasking multivariate outliers and leverage points (with discussion). *J. Amer. Statist. Assoc*, 85:633–651, 1990.

P. J. Rousseeuw. Multivariate estimation with high breakdown point. *Mathematical statistics and applications*, 8:283–297, 1985.

P. J. Rousseeuw and C. Croux. Alternatives to the median absolute deviation. *Journal of the American Statistical association*, 88(424):1273–1283, 1993.

P. J. Rousseeuw and K. V. Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3):212–223, 1999.

R. Serfling. A depth function and a scale curve based on spatial quantiles. In *Statistical Data Analysis Based on the L1-Norm and Related Methods*, pages 25–38. Springer, 2002.

C. G. Small. A survey of multidimensional medians. *International Statistical Review/Revue Internationale de Statistique*, pages 263–277, 1990.

M. Sokolova, N. Japkowicz, and S. Szpakowicz. Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation. In *Australasian Joint Conference on Artificial Intelligence*, pages 1015–1021. Springer, 2006.

Y. Sun and M. G. Genton. Functional boxplots. *Journal of Computational and Graphical Statistics*, 20(2):316–334, 2011.

G. Tarr, S. Müller, and N. C. Weber. Robust estimation of precision matrices under cellwise contamination. *Computational Statistics & Data Analysis*, 93: 404–420, 2016.

M. Templ, P. Filzmoser, and C. Reimann. Cluster analysis applied to regional geochemical data: problems and possibilities. *Applied Geochemistry*, 23(8):2198–2213, 2008.

J. W. Tukey. Mathematics and the picturing of data. In *Proceedings of the international congress of mathematicians*, volume 2, pages 523–531, 1975.

Y. Vardi and C.-H. Zhang. The multivariate l1-median and associated data depth. *Proceedings of the National Academy of Sciences*, 97(4):1423–1426, 2000.

J. A. Vargas N. Robust estimation in multivariate control charts for individual observations. *Journal of Quality Technology*, 35(4):367–376, 2003.

S. Verboven and M. Hubert. Libra: a matlab library for robust analysis. *Chemometrics and intelligent laboratory systems*, 75(2):127–136, 2005.

E. J. Wegman. Hyperdimensional data analysis using parallel coordinates. *Journal of the American Statistical Association*, 85(411):664–675, 1990.

Y. Zeng, G. Wang, E. Yang, G. Ji, C. L. Brinkmeyer-Langford, and J. J. Cai. Aberrant gene expression in humans. *PLoS Genet*, 11(1):e1004942, 2015.